

C 012- H0338-3- 6335077

AMERICAN PHILOSOPHICAL QUARTERLY

3

Edited by
NICHOLAS RESCHER

With the advice and assistance of the Board of Editorial Consultants:

William Alston
Alan R. Anderson
Kurt Baier
Lewis W. Beck
Richard B. Brandt
Roderick M. Chisholm
L. Jonathan Cohen
James Collins
Michael Dummett

James M. Edie
Peter Thomas Geach
Adolf Grünbaum
Carl G. Hempel
Jaakko Hintikka
Raymond Klibansky
Benson Mates
John A. Passmore
Günther Patzig

Richard H. Popkin
Wesley C. Salmon
George A. Schrader
Wilfrid Sellars
J. J. C. Smart
Wolfgang Stegmüller
Manley H. Thompson, Jr.
G. H. von Wright
John W. Yolton

VOLUME 3/NUMBER 1

CONTENTS

JANUARY 1966

- | | | | |
|---|----|---|----|
| I. K.W. RANKIN: <i>Wittgenstein on Meaning, Understanding, and Intending</i> | I | VI. G. B. KEENE: <i>Can Commands Have Logical Consequences?</i> | 57 |
| II. JOHN W. YOLTON: <i>Agent Causality</i> | 14 | VII. J. W. MEILAND: <i>Temporal Parts and Spatio-Temporal Analogies</i> | 64 |
| III. FREDERIC B. FITCH: <i>Natural Deduction Rules for Obligation</i> | 27 | VIII. HOWARD SMOKLER: <i>Goodman's Paradox and the Problem of Rules of Acceptance</i> | 71 |
| IV. HERBERT HOCHBERG: <i>Things and Descriptions</i> | 39 | IX. HUGH S. CHANDLER: <i>Three Kinds of Classes</i> | 77 |
| V. PETER UNGER: <i>On Experience and the Development of the Understanding</i> | 48 | X. WILLIAM H. CAPTAN: <i>Part X of Hume's Dialogues</i> | 82 |

PUBLISHED BY BASIL BLACKWELL WITH THE COOPERATION OF THE UNIVERSITY OF PITTSBURGH

AMERICAN PHILOSOPHICAL QUARTERLY

POLICY

The *American Philosophical Quarterly* welcomes articles from philosophers of any country on any aspect of philosophy, substantive or historical. However, only serious and original articles will be published, and not news items, book reviews, critical notices, or "discussion notes."

MANUSCRIPTS

Contributions may be as short as 2,000 words or as long as 25,000. All manuscripts should be typewritten with wide margins, and at least double spacing between the lines. Footnotes should be used sparingly and should be numbered consecutively. They should also be typed with wide margins and double spacing. The original copy, not a carbon, should be submitted; authors should always retain at least one copy of their articles.

EDITORIAL COMMUNICATIONS

Articles for publication, and all other editorial communications and enquiries, should be addressed to: The Editor, *American Philosophical Quarterly*, Department of Philosophy, University of Pittsburgh, Pittsburgh, Pennsylvania 15213.

OFFPRINTS

Authors will receive gratis two copies of the issue containing their contribution. Offprints can be purchased at low cost through arrangements made when checking proof.

SUBSCRIPTIONS

The price *per annum* is six dollars for individual subscribers and ten dollars for institutions. Checks and money orders should be made payable to the *American Philosophical Quarterly*. Back issues are sold at the rate of two dollars to individuals, and three dollars to institutions. All correspondence regarding subscriptions and back orders should be addressed directly to the publisher (Basil Blackwell, Broad Street, Oxford, England).

* * *

335077



AMERICAN PHILOSOPHICAL QUARTERLY

Edited by
NICHOLAS RESCHER

With the advice and assistance of the Board of Editorial Consultants:

William Alston
Alan R. Anderson
Kurt Baier
Lewis W. Beck
Roderick M. Chisholm
L. Jonathan Cohen
James Collins
James M. Edie
Peter Thomas Geach

Adolf Grünbaum
Carl G. Hempel
John Hospers
Raymond Klibansky
Ernan McMullin, S.J.
Benson Mates
John A. Passmore
Günther Patzig
Richard H. Popkin

Wesley C. Salmon
George A. Schrader
Wilfrid Sellars
Alexander Sesonske
J. J. C. Smart
Manley H. Thompson, Jr.
James F. Thomson
G. H. von Wright
John W. Yolton



VOLUME 3 (1966)

PUBLISHED BY BASIL BLACKWELL WITH THE COOPERATION OF THE UNIVERSITY OF PITTSBURGH

AMERICAN PHILOSOPHICAL QUARTERLY

CONTENTS OF VOLUME 3 (1966)

	Page
ABELSON, RAZIEL <i>Persons, P-Predicates, and Robots</i>	306
ALLEN, DIOGENES <i>Motives, Rationales, and Religious Beliefs</i>	111
BAIER, KURT <i>Moral Obligation</i>	210
BAKER, G. P., AND P. M. HACKER <i>Rules, Definitions, and the Naturalistic Fallacy</i>	299
CAPTAN, WILLIAM H. <i>Part X of Hume's Dialogues</i>	82
CHANDLER, HUGH S. <i>Three Kinds of Classes</i>	77
CHISHOLM, RODERICK M., AND ERNEST SOSA <i>On the Logic of "Intrinsically Better"</i>	244
CRITTENDEN, CHARLES <i>Fictional Existence</i>	317
EISENBERG, PAUL D. <i>Basic Ethical Categories in Kant's Tugendlehre</i>	255
FEINBERG, JOEL <i>Duties, Rights, and Claims</i>	137
FITCH, FREDERIC B. <i>Natural Deduction Rules for Obligation</i>	27
GALE, RICHARD M. <i>McTaggart's Analysis of Time</i>	145
HACKER, P. M., <i>see</i> G. P. BAKER AND P. M. HACKER	
HOCHBERG, HERBERT <i>Things and Descriptions</i>	39
KEENE, G. B. <i>Can Commands Have Logical Consequences?</i>	57
KIM, JAEGWON <i>On the Psycho-Physical Identity Theory</i>	227
MCCALL, STORRS <i>Temporal Flux</i>	270
MACRAE, VALERIE, <i>see</i> RICHARD ROUTLEY AND VALERIE MACRAE	
MEILAND, J. W. <i>Temporal Parts and Spatio-Temporal Analogies</i>	64
OWENS, JOSEPH <i>The Grounds of Universality in Aristotle</i>	162
PERKINS, MORELAND <i>Emotion and the Concept of Behavior</i>	291
PRIOR, A. N. <i>Postulates for Tense-Logic</i>	153
PURTILL, R. L. <i>Moore's Modal Argument</i>	236
RANKIN, K. W. <i>Wittgenstein on Meaning, Understanding, and Intending</i>	1
ROUTLEY, RICHARD AND VALERIE MACRAE <i>On the Identity of Sensations and Physiological Occurrences</i>	87
SCOTT-TAGGART, M. J. <i>Recent Work on the Philosophy of Kant</i>	171
SESONSKE, ALEXANDER <i>Moral Rules and the Generalization Argument</i>	282
SIEGLER, FREDERICK A. <i>Lying</i>	128
SMOKLER, HOWARD <i>Goodman's Paradox and the Problem of Rules of Acceptance</i>	71
SOSA, ERNEST, <i>see</i> RODERICK M. CHISHOLM AND ERNEST SOSA	
UNGER, PETER <i>On Experience and the Development of the Understanding</i>	48
VAN DE VATE, DWIGHT, JR. <i>Other Minds and the Uses of Language</i>	250
WALLACE, JAMES D. <i>Pleasure as an End of Action</i>	312
YOLTON, JOHN W. <i>Agent Causality</i>	14

I. WITTGENSTEIN ON MEANING, UNDERSTANDING, AND INTENDING

K. W. RANKIN¹

I. MEANING AS THE SOCIAL SYNTAX OF LANGUAGE BEHAVIOR

IN his comments on philosophical misuse of the word "use," Ryle² has made at least³ two implicit criticisms of the Wittgensteinian analysis of meaning as use.⁴ As he sees it the word can apply to a way of operating with words or other devices which are subsidiary to sayings, but not to the sayings themselves. Hence one might say, though he does not, that if we use the word to talk about language, it can only refer to the internal syntax of sayings. His most crucial point, however, is that "use" does not mean "usage" except in archaic English. A usage is a custom, practice, fashion, or vogue, not a way of, or technique for, operating with something. There is an opposition between use and misuse, but there is no such thing as a misuse. Ryle holds that descriptions of usages presuppose descriptions of use, though he qualifies this somewhat by adding that in interpersonal transactions learning to use their instruments involves noticing how other people operate these instruments.

One suspects, on the other hand, that German words such as "Gebrauch" and "Verwendung" might, even in contemporary use, lend themselves more readily to the sort of conceptual structure in which no true Englishman could feel at ease. Certainly Wittgenstein, both in his own English and in that of his translators, shows no compunction about using "use" and "usage" interchangeably, or about applying either of these words to

sayings as well as to words. One must hasten to add, however, that this is no mere linguistic idiosyncrasy either of an individual or language. So far as I can see, Wittgenstein has quite explicitly rejected the conceptual distinctions in terms of which Ryle has restricted the application of these words. In fact, this conceptual assimilation could reasonably be described as Wittgenstein's major preoccupation.

Very early in the *Blue Book*⁵ he maintains that what gives the signs in an expression their life is just their use. The penultimate sentence of the same work takes the point further by enjoining us not to "imagine the meaning (*sc.* of an expression) as an occult connection the mind makes between a word and a thing, and that this connection contains the whole usage of a word as the seed might be said to contain the tree." The tail of this injunction acquires an even more pointed sting when it reappears in the *Brown Book*⁶ in the form of the complaint that "We meet again and again with this curious superstition, as one might be inclined to call it, that the mental act is capable of crossing a bridge before we've got to it." To go by Moore's report,⁷ Wittgenstein's thoughts were already taking this direction in 1930-31, and passages parallel to these proliferate in the *Investigations*, where, of course, the argument is developed in greater dialectical detail.

The *Investigations* would seem quite explicitly to reject Ryle's conceptual distinction between way of, or technique for, operating and custom in passages such as the following:

¹ An emended version of a Presidential Address to the Victorian Branch of The Australasian Association of Philosophy delivered in March, 1964.

² Gilbert Ryle, "Ordinary Language," *Philosophical Review*, vol. 62 (1953), pp. 167-186.

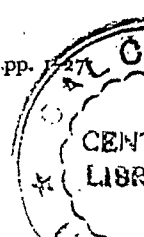
³ Wittgenstein describes words and descriptions as tools or instruments (*Investigations*, I, §§ 11, 23, 291), but since tools are cultural objects, i.e., things imbued by custom and usage, I take it he is not relying upon a direct equation between "use" and "utility" of the sort Ryle also condemns.

⁴ *Investigations*, I, §§ 30, 43, 138. References to particular passages in *The Blue and Brown Books* (Oxford, Blackwell, 1958) and to *The Philosophical Investigations* (Oxford, Blackwell, 1953) are to part and section except where otherwise indicated.

⁵ *Blue Book*, p. 4.

⁶ *Brown Book*, II, 5, p. 143.

⁷ G. E. Moore, "Wittgenstein's Lectures 1930-33," *Mind*, vol. 63 (1954), pp. 1-15, 289-316 and vol. 64 (1955), pp. 1-27, reprinted in *Moore's Philosophical Papers* (London, Allen and Unwin, 1959), see especially p. 259.



It is not possible that there should have been only one occasion on which someone obeyed a rule. It is not possible that there should have been only one occasion on which a report was made, an order given or understood; and so on—To obey a rule, to make a report, to give an order, to play a game of chess are *customs* (uses, institutions). To understand a sentence means to understand a language. To understand a language means to be master of a technique.⁸

This passage also makes clear how its conceptual assimilation of use and custom brings it about that both concepts apply to sayings, i.e., speech activities such as reportings and commandings, as well as to words. And in other passages Wittgenstein constructs language-games within which the distinction between word and saying doesn't even properly arise.⁹

Again, it becomes clear in the passage just before the above quotation that Wittgenstein has forearmed himself against Ryle's argument that descriptions of usage presuppose descriptions of uses and techniques. To illustrate his point Ryle had observed that Mrs. Beeton's recipe for omelets mentions or assumes nothing about Parisian chefs, whereas Baedeker's guide to the latter's customs must, if complete, mention or presume familiarity with the recipe. For Wittgenstein this distinction rests implicitly on the sort of assumption to which both he and surely the author of *The Concept of Mind* as well as explicitly opposed. The assumption would seem to be that correct usage requires as its source some inner act of meaning or understanding whereby our speech, or response to speech, is guided by privately consulted prescriptions for the correct use of language. The distinction between technique and usage would seem to model itself upon the distinction between the perforated scroll tucked inside the pianola and the sounds which the pianola emits, with the further consequence that understanding the use is modeled upon the internal mechanical processes whereby the instrument scans the

perforations and translates the pattern into sounds.¹⁰

As a corrective he now¹¹ distinguishes between (a) the causal question how the expression of a rule, e.g., a signpost, leads to action and (b) the entirely different sort of question of what going by a rule, e.g., following a signpost is. The cause consists just in the training, or possibly even in a natural propensity,¹² whereas the actual going by the rule consists just in the custom for a regular use of signposts. Now the tendency to think of the technique or way of operating with a sign as something which we grasp by means of an inner mental process, and the tendency to distinguish between the technique and the custom or usage, play into each other's hands, or are two facets of the same mistake, according to Wittgenstein. Both result from a confusion between the causal and the non-causal questions.

To illustrate more effectively how a theory of meaning, and the philosophical psychology of meaning, might be conceived as profiting from this distinction, I now propose to follow Wittgenstein's own lead by investigating an anthropological model. The model I propose, however, does not consist directly in the way of life of some primitive tribe, i.e., the interwoven patterns of language and action which he called language-games. More accurately my model is meta-anthropological, viz., a current controversy among anthropologists about techniques of investigation.¹³ Perhaps this will lead to caricature, but at least it should bring the relevant issues into sharper focus.

Associates of Evans-Pritchard¹⁴ in England and Lévi-Strauss in France stress the danger of examining any particular social institution except in relation to the structure of a particular society as a whole. We might describe their practice in Wittgenstein's phrase as "descriptive analysis." Other anthropologists, and specifically Homans and Schneider in a recent book,¹⁵ seek to explain

⁸ *Investigations*, I, § 199.

⁹ *Brown Book*, I, 1; *Investigations*, I, § 19.

¹⁰ Cf. *Brown Book*, I, 66.

¹¹ *Investigations*, I, § 198.

¹² Cf. *Brown Book*, I, 40.

¹³ See Rodney Needham's monograph on method, *Structure and Sentiment* (Chicago, University of Chicago Press, 1962).

¹⁴ Evans-Pritchard himself may in some ways appear atypical of this group. While insisting upon the holistic approach to the study of any community, at the same time he places greater emphasis than others upon an historical understanding (see his *Essays in Social Anthropology* [London, Faber & Faber, 1962], pp. 18 ff.). Apparently he wishes to combat the tendency toward quasi-biological determinism which he finds in Malinowski. Some of his associates prefer to preserve the autonomy of their discipline and its holistic methods by confining its scope to non-transitional phases in stable societies.

¹⁵ *Marriage, Authority and Final Causes: A Study of Unilateral Cross-Cousin Marriage* (Glencoe, Illinois, Free Press, 1955).

such phenomena by reference to individual sentiment. A test case of the two methods can be found in their application to the institution of prescriptive unilateral cross-cousin marriage. To put it very crudely, this institution requires each male in a tribe to marry only from cousins on his mother's side, and forbids him to marry from cousins on his father's side, or vice-versa.

Now the "descriptive analysts" show how the matrilinear form of cross-cousin marriage (i.e., marriage on the mother's side) brings within each generation a circular movement of women among the families so linked, and how this movement is counterbalanced by a reciprocating circular movement of chickens, pigs, services, labor, etc., in the reverse direction. These reciprocating movements are basic to the whole intricate pattern of the life of the tribe. The structure of the dwellings, the organization of the sleeping quarters and the living area, the ritual practices and the parts in them played by various members of the family, the status which individuals accord to each other, and a host of other features, which might otherwise seem unrelated and arbitrary, all fall into place once one understands the marriage prescriptions.

With one eye on our philosophical analogue we might observe that in this way social institutions possess what Wittgenstein, in another connection, called "depth." From an external point of view their prescriptions are arbitrary, but to the individual born into the society its incest laws will probably seem to have, at least initially, an essential and unquestionable rightness, since he is already, and quite unthinkingly, committed to their observance throughout a wide range of his activities. And it would be difficult for him to consider questioning them by *active* nonconformity unless he happened to be already in other respects partially detribalized.

Notice too that in so far as the descriptive analysts make any attempt to *explain* why one type of incest law rather than any other should prevail in a given society, it is in terms of what they call the "solidary" consequences. With the aid of schematized genealogical charts representing different possible patterns of interrelationship between families they show how, say, prescriptive matrilinear cross-cousin marriages must lead to a greater degree of socio-economic integration than the prescriptive patrilinear sort. In the matrilinear case the reciprocating circulations of women and gifts will continue in the same opposing directions from generation to generation,

while in the patrilinear case they must change direction from generation to generation with disrupting social consequences. Hence the matrilinear form has greater persistence-value. This deduction can then be checked empirically by showing that few or no societies have been observed to practice prescriptive patrilinear cross-cousin marriages of an unambivalent kind. But it is important to bear in mind that at the same time descriptive analysts are not at all anxious to engage in explanation. Their primary concern is to understand the structure of particular societies, rather than to devise theories which have a general explanatory force.

Homans and Schneider, on the other hand, have tried to *explain* the practice of prescriptive unilateral cross-cousin marriages of either form by a causal type of hypothesis, of which the following is a simplification. When descent is from the father's side (i.e., patrilineal), a prescriptive unilateral cross-cousin marriage will be matrilinear, whereas when descent is from the mother's side (i.e., matrilineal) it will be patrilinear. The reason offered is that in the patrilineal case feelings of constraint will exist between father and son, since parental authority rests with the father, and a corresponding indulgent relation will exist between mother and son. These sentiments will permeate respectively the attitudes of the son and his father's relations toward each other, on the one hand, and the attitudes of the son and his mother's relations toward each other, on the other. Hence in the patrilineal society, matrilinear marriage will be sentimentally the more appropriate; and for similar reasons, *mutatis mutandis* in the matrilineal society, patrilinear marriage will be sentimentally the more appropriate.

The respective merits of these two different methods of anthropological investigation don't, of course, concern us here. Our concern is with the points of contact between this issue and the philosophical issues which Wittgenstein has raised. Five points of comparison are quite striking.

(1) The descriptive anthropological school might be interpreted as holding that the "meaning" of a social form of behavior lies in its social syntax, i.e., in the functional interdependence between it and different forms of behavior within a spatio-temporal extended area. Similarly, but in a more literal sense of "meaning," Wittgenstein would appear to hold that the meaning of language-behavior lies in its interdependence with other forms of behavior within the total way of life of

the community which plays the language-game in question.¹⁶

(2) The former condemn the psychologicistic methods of some of their contemporaries as just extensions of the less sophisticated scissors-and-paste technique of older generation writers such as Frazer in *The Golden Bough*. Similarly Wittgenstein describes the more traditional type of philosopher as having been bewitched by language while it was on holiday or idling.¹⁷

(3) The former fight shy of causal explanation of social behavior in terms, say, of individual sentiment. Similarly Wittgenstein rejects causal explanation of linguistic behavior in general in terms of mental states or processes as beside the point.¹⁸ The meaningfulness consists just in the social syntax.

(4) The former treat social prescriptions as arbitrary, but yet as possessing a not easily questioned authority through their mutual integration and solidary consequences. Similarly Wittgenstein comments on the depth of the grammar of linguistic behavior.¹⁹

(5) The former treat social prescriptions as arbitrary, but yet as limited by certain very general facts about human nature²⁰ and its environment, which determine the persistence-value of these institutions. Similarly Wittgenstein regards grammatical rules as arbitrary, but yet as limited by certain very general facts of nature.²¹

II. MEANING AND UNDERSTANDING

The comparison has, I think, been sufficiently particular and many-sided to establish the meta-anthropological model as at least pertinent. But, unless its use constitutes an utter travesty, it indicates very directly that Wittgenstein's concept of meaning is insufficiently discriminate. The meaning which the anthropologist finds in social behavior as such is not related to understanding in the way in which the meaning which we ascribe to language behavior as such is related to understanding, and this can be seen in a number of ways.

(1) The sort of meaning which consists merely of functional interdependence need not go with any form of understanding at all. It is displayed, for instance, by the economy of nature, or the balance of life, or the organs within an organism, and the presence of something which understands is only essential to specific forms of these phenomena.

(2) The functional interdependence displayed by the social behavior of specifically human communities does display a corporate know-how or dispositional form of understanding. Those who participate must have mastered certain techniques of stimulus and response, though, of course, not in all cases the same techniques. But the most one could say of these forms of understanding is that conjointly they constitute the functional interdependence of the social behavior. They must not be confused with an understanding of the functional interdependence, far less with an understanding of the sort of meaning which is specific to language. It is the anthropologist alone, or the comparatively rare individual who reflects on the ways of his own community, who understands, or has an understanding of, the functional interdependence. Now, by way of contrast, to understand a language certainly involves the mastery of a technique or an ability, but the ability is partly that of being able to understand, or have an understanding of, what is said to one on specific occasions at least some of the time. Consequently, since this understanding of is essential to the language know-how, but not to the sort of know-hows or dispositional forms of understanding which merely constitute functional interdependence, it follows that the meaning specific to language cannot consist simply in functional interdependence. This is not to deny that the functional interdependence between language and other forms of behavior is an important part of the former's significance. But this is the kind of significance which Austin investigated under the more circumspect heading of "force."²²

¹⁶ *Investigations*, I, §§ 7, 20, 23.

¹⁷ *Ibid.*, I, §§ 38, 83, 109, 132.

¹⁸ *Ibid.*, I, §§ 109, 371, 497.

¹⁹ I owe this point to Mr. L. B. Grant. As further evidence for the need to introduce a distinction between different modes of significance, note that the words in our language do not function altogether in the manner of cards in a pack. The cards tend to change their value from one sort of game to another, but to a considerable extent the possibility of "playing" different "games" with words depends upon their meaning remaining constant (cf. R. Rhees: "Wittgensteins' Builders," *Proceedings of the Aristotelian Society*, vol. 60 [1959-60], pp. 179-180). Consider the use of "Your wife and family are still behind the Iron Curtain" used on different occasions by the appropriate person or persons in the course of deliberating, informing, misinforming, advising, warning, threatening, questioning, etc. But note that Wittgenstein explicitly repudiates the early Fregean version of the Austin-Hare sort of distinctions which these examples may seem to suggest. (*Investigations*, I, § 22.)

²⁰ *Ibid.*, I, §§ 89, 111.

²¹ See Needham, *op. cit.*, pp. 27-28.

²² *Investigations*, I, §§ 56, 372, 496; II, xii.

(3) Finally, if meaning consists of functional interdependence *every* institution in a community would have to be part of that community's language, and not just what normally goes by that name. Prescriptive matrilinear cross-cousin marriages would have the same sort of meaning as a word or sentence. But it isn't just that the tongue is *more* eloquent than the sexual organs. As the organ of speech its eloquence is of a quite different kind.

For these three reasons, then, we must conclude that the type of meaning peculiar to usages, by virtue of their functional interdependence, is not related to understanding in the way that the meaning peculiar to language is related to the understanding of that meaning. But if this is so, we have now to come to terms with Wittgenstein's arguments. To confront his conclusions with commonsensical reminders is not enough. Should his arguments lead us to revise these reminders? More specifically, do they show that unless we interpret meaning as functional interdependence, and modify our concept of understanding accordingly, we are forced into conceiving understanding as an occult sort of phenomenon which explains correct usage? Or alternatively, do they show that to conceive understanding in this way as an explanatory phenomenon is mistaken?

I shall give a negative answer to the first of these questions, and an affirmative to the second. In language meaning is primarily not functional interdependence, and understanding not some occult source of correct behavior. And to uphold this I wish now to sketch more fully, but still in a preliminary sort of way, the set of distinctions between different sorts of understanding at which my preceding arguments have already hinted.

First we must distinguish a dispositional form of understanding, the sort which we also call "a know-how." This sort seems to be subdividable into at least two kinds. One species is both realized in and displayed by "overt" behavior. Such things as the ability to play golf, or to speak a language or to write it, belong to this kind. The display of such skills may be functionally interdependent with the display of other skills by other people or oneself, and hence may be a constituent of a way of life which could be said to realize a corporate form of understanding or know-how.

The other species consists of skills which are displayed by, but not realized in, "overt" be-

havior. They are realized by a different non-dispositional or episodic form of understanding. Such things as hearing, seeing, understanding the meaning of something written or printed in a particular place, or quite generally understanding a particular situation, are all examples of the episodic form of understanding which realizes, but does not display, skills belonging to the second species of dispositional understanding. But note in passing that something that often goes under the name of the one skill, e.g., the understanding of a language, may in fact involve skills belonging to both dispositional kinds, e.g., the skill to speak and the skill to understand what is spoken.

Now the distinction between the two types of dispositional understanding cannot be eliminated by reducing what I have called the episodic form to a dispositional form of the first type for two reasons.

(1) As P. T. Geach points out,²³ whether one's understanding of a particular situation is displayed or not depends upon one's desires as well as upon one's understanding. Hence this sort of understanding is not the sort of thing which is realized in a display.

(2) The behavior which both realizes and displays a dispositional form of understanding must be intentional, for this is what differentiates such displays as a display of a know-how or skill, as distinct from a display of some other sort of disposition. I don't mean, of course, that the displayer must form an explicit intention to behave as he does, but merely that the display must not conflict with any of his intentions, and that the fact that it doesn't so conflict is an indispensable condition of the performance. Now intentional behavior presupposes an understanding by the agent of the situation in which he is called upon to act. The golfer's display of skill, for instance, is only considered to be such if we suppose him to have understood the lay of the land, the possible influence of the wind, the properties of his clubs, etc., on particular occasions, though no doubt this understanding cannot be divorced altogether from the muscular finesse which he has acquired through training. Consequently, the skill or knack which consists in understanding particular sorts of situation on different occasions cannot be reduced to the first sort of skill since it is presupposed by it. We have then to distinguish between skills or knacks which are both realized in and displayed by intentional behavior and those which are

²³ *Mental Acts* (London, Routledge and Kegan Paul, 1957).

realized in something which, strictly speaking, can be neither intentional nor unintentional, since it is presupposed by intentional behavior.

These distinctions, no doubt, merit independent consideration, but in what follows I shall simply show that the encounter with Wittgenstein's analysis of understanding leads to their refinement rather than their destruction. He tries to show (a) that specific episodic states such as imagining, private consultations of rules, or the presence of some sort of atmosphere, cannot possibly be essential to the understanding of language, and (b) more generally, that to represent understanding, or meaning, as any sort of episodic state is to misunderstand the grammar of such words as "understand," "mean," and "intend." I shall confine myself, however, to the more general claim,²⁴ and now comment in turn upon successive stages in the following exposition of his supporting arguments.

III. UNDERSTANDING NOT A MENTAL STATE OR PROCESS

Exposition—The grammar of such words or phrases as "knows," "can," "is able to," "understands," and "I meant"²⁵ is very much alike and closely related to the notion of the mastery of a technique. It is markedly unlike that of words for mental states such as "depression," "excitement," "pain," etc. One can readily interpret somebody's claim to have felt pain continuously since yesterday. But how would one construe the claim to have understood a specific word continuously? Similarly, we wouldn't normally ask such questions as: "When do you know how to play chess? All the time? Or just while you are making a move? And the whole during each move?"²⁶

On the surface, however, words like "understand" may sometimes display the same grammar as words like "pain" in the following way. Wittgenstein goes on to note:

But there is also *this* use of the word "to know": we say "Now I know!"—and similarly "Now I can do it!" and "Now I understand!"²⁷

The meaning or the whole use of a word, or a

formula, or procedure for expanding, say, an arithmetical series seems to occur to one in a flash.²⁸ This similarity, however, must not lead one to ignore the differences, and Wittgenstein seeks to place both in perspective in two somewhat unrelated ways.

He draws an illuminating analogy between the group of expressions like "understand" and the word "fit" as used to describe the relation between two hollow cylinders *A* and *B*.²⁹ The point can be summarized, though rather crudely, as follows. The word "fit," like those other expressions, indicates a disposition rather than a state. *A* fits into *B*, even when not stuck into *B*. There are, in fact, a number of criteria for saying that *A* fits *B* besides *A* being stuck into *B*; but to say it fits is not to refer to any such state or collection of states. Finally, we can say that *A* has begun, or ceased, to fit *B*; but, again, we are not referring to the specific episode or episodes which may have served as criteria for the statement.

This, then, suggests how the word "understand" might be used in a seemingly episodic way without actually referring to any specific episode such as the occurrence of a formula, a feeling of relaxation, etc. Without the aid of this analogy we might consequently have been led to believe that, since it didn't refer to any specific assignable episode, it must refer to some hidden or occult episode.³⁰

Wittgenstein, however, develops an additional account of the grammar of "Now I understand!" which attaches significance to the fact, not merely that it is in the present or, in any case, that it refers to a definite date, but also to the fact that it is in the first person. He now contrasts the point of view of the speaker with that of the audience. The speaker may say it when the formula occurs to him, or when he experiences a feeling of relief, or even when he experiences nothing at all except the inclination to say it, and we may underwrite what he says given certain circumstances, viz., that he knows algebra, or that he has always proved reliable in the past. But although these particular circumstances constitute *our* justification and reason for underwriting what he says, they would not normally constitute the reason why he

²⁴ In a much longer version of this paper I satisfied myself that Wittgenstein's attempt in (a) really presupposed the more general and positive claims leading him to (b), and also that any distinction between two forms of understanding need not identify the episodic form exclusively with any of the episodic states specified in (a). The lengthiness of the argument eventually compelled me to presume upon its results.

²⁵ *Investigations*, I, §§ 150, 187.

²⁶ *Ibid.*, § 59.

²⁷ *Ibid.*, I, § 151; cf. I, § 179.

²⁸ *Ibid.*, I, §§ 138, 139, 197.

²⁹ *Ibid.*, I, § 182; cf. *Brown Book*, I, 44.

³⁰ *Investigations*, I, § 153.

says it. His claim is not made on the basis of past experience.³¹

This again might tempt one to suppose that the understanding actually consists in some mysteriously hidden experience accessible only to the speaker. But while Wittgenstein does not seem to wish to deny that there are in some sense experiences of this sort, he cannot allow them to be described as states of understanding. Whether the speaker has understood or not depends upon the development of his future behavior. Understanding cannot be conceived of as a state which is the source of correct use,³² since failure to evince the correct behavior in the future must count, even for the speaker, as at least very strong *prima facie* evidence against his earlier claim that he understood.

Consequently Wittgenstein places this use of "Now I understand!" in the category of what he calls "a signal" ("Signal").³³ And in other contexts he appears to use such words as "exclamation" ("Ausruf"),³⁴ "cry" ("Schrei"),³⁵ "manifestation" ("Äusserung"),³⁶ and "response" ("Reaktion")³⁷ to fulfill the same sort of function, i.e., to indicate that their use is at least not solely to make reports. Instead of looking for some state of affairs which they describe as an explanation of their use we should look simply at what is said as a proto-phenomenon. We must be contented to say "*This is how these words are used*" or "*This language-game is played*."³⁸ If we follow out this recommendation, we come to see that the function of such words or sentences is to draw attention to the utterer rather than, or as well as, to make a report.

Comment—In his analysis of understanding as the mastery of a technique, Wittgenstein seems to be bringing understanding under the category of a disposition. Now his destructive criticism of attempts to find the cause of this mastery in some mental and introspectible state seems valid to me. But the criticism would seem only to affect the dispositional form of understanding which is both realized in and displayed by activities such as speech activities which proceed in accordance with formation rules. It doesn't affect the episodic form of understanding in which the listener's mastery of a language is realized on particular occasions of speech, for this episodic form is not being

invoked as the source of some correct procedure. To postulate it is not to hypostatize something hidden which corresponds to an ability, for *ex hypothesi* it has been postulated simply as that in which the ability has been realized.

No doubt I have drawn the distinction between the speaker's ability and the listener's somewhat too sharply. On the one hand, it is usually intentional on the part of a speaker who has mastered a language that the listener who has mastered the language should understand what is said in the way that he does. On the other, the listener's mental state only acquires the status of understanding by virtue of his ability to display the speaker's ability, or to respond correctly to what is said in other ways.

But Wittgenstein doesn't seem to draw the distinction sharply enough. I think the reason for this must be, that if you attribute the meaning of a language to the functional interdependence between the *actions* of speakers and listeners, then the dispositional sort of understanding, which both types of action display, will in each case have to be the sort of ability which is realized in that which displays it. If I am right in this, then this part of Wittgenstein's argument must in fact presuppose the conception of language which I have been requiring it to support.

Perhaps the most interesting part of Wittgenstein's argument lies in his distinction between first and other person uses of the word "understand." This, however, can more conveniently be considered under the subject of intention which I raise in Section V. For the moment it is enough to note that the sort of episode with which his first person use is concerned is the *acquiring* of a dispositional understanding, e.g., the speaker's ability, and not the episodic understanding which one now has of a particular situation, e.g., a particular utterance.

IV. UNDERSTANDING AS ACHIEVEMENT

Exposition—As an auxiliary to his distinction between the first and other person use of the word "understand" Wittgenstein makes use of a somewhat ambivalent analogy between understanding and reading.³⁹ Reading is here taken in its minimal

³¹ *Ibid.*, I, §§ 147, 179.

³² *Ibid.*, I, § 146.

³³ *Ibid.*, I, § 180.

³⁴ *Ibid.*, I, 323; II, xi, pp. 197, 218.

³⁵ *Ibid.*, II, xi, p. 197.

³⁶ *Ibid.*, I, § 585.

³⁷ *Ibid.*, I, § 659.

³⁸ *Ibid.*, I, §§ 180, 654.

³⁹ *Ibid.*, I, §§ 156-164.

form, as simply the translation of writing, or print, into sounds. Now it might be held that this process of translation must involve some sort of private mental correlation between written shapes and sound in accordance with a rule. Against this Wittgenstein argues as follows.

(a) We are inclined to this view when we think of (say) reading by a beginner. But reading covers a large family of activities, of which the beginner's is only one case. One's eye may pass along a printed line letter by letter, syllable by syllable, or whole words or phrases at a time. One may speak neither aloud, nor to oneself, at the time, but be able to repeat what one had read later on. One may read attentively, or inattentively, as if one were a mere reading machine, and so on.

(b) Is there any marked difference between the experience of the learner who first learns a passage by heart, and then pretends to read it by looking at the script, and the person who reads effortlessly, but inattentively?

(c) Again, a beginner may read at first with only spasmodic success, making it difficult to decide whether he hits on the right words by accident or not. But then by a gradual transition he may come to read with complete accuracy. Now does it make sense to ask which was the first word he really read?

(d) Suppose for a moment that true reading were distinguishable from false reading by the presence of some conscious act or characteristic sensation. Now couldn't this feeling be induced or inhibited by some drug so as to be present, as sometimes in dreams when one is faced by unintelligible signs, or absent where the signs *are* intelligible.

(e) To test whether there must be some conscious act which transforms the scanning process into reading try the following experiment. Say the numbers from 1 to 12, and then look at the dial of your watch and *read* them. Can you really detect some process additional to the scanning that turns the repetition into reading?

Perhaps one might risk using Rylean terms to express Wittgenstein's main conclusion from this battery of arguments. One might say that "reading" is really an achievement word. It is used when any of an indefinite variety of processes meets certain criteria of success, irrespective of whether these activities are performed laboriously, deliberately, automatically, etc. If we fail to realize this, however, we may feel that the phenomenon of reading isn't fully present in all its various manifestations, but must also involve something

ghostly or ethereal. Similarly, to say on the basis of someone's handling of a language that he understands it is not to point to some inner process accompanying the actual handling, even if it is not simply to describe the way in which he handles it.

Comment—I think that this use of the analogy owes a good measure of its apparent destructive force to the fact that there are at least two ways in which reading might seem comparable to understanding, and two ways in which the use of reading as a model might seem to lead to the postulation of understanding as a mental episode. Of these two ways I think that one is mistaken for the reasons that Wittgenstein has given, but that these reasons don't affect the second way at all. And I also think that it is in the second way, the one unaffected by Wittgenstein's reasons, that the analogy between reading and understanding is more compelling. As a consequence of this, his arguments against episodic understanding may seem to have a greater destructive force than they actually exert if we fail to notice, as we may, that they affect validly only the less obvious way in which the model of reading seems to suggest an episodic understanding.

The less obvious way would take the beginner's case as typical, and then suggest that in the same way as the beginner may in reading make mental use of rules for translating letters into sounds, so, when displaying our mastery of a language, we must be having private recourse of some sort to formation-rules. And here Wittgenstein's argument that the beginner's case is not typical, and that what makes reading "reading" is the success of his performance, does seem to point to the fact that what makes the understanding, i.e., the mastery, of the language a mastery, is once more the success of the operation with words, not some private invoking of formation rules.

But the second way in which understanding might seem comparable to reading is in the end-product of the latter. Just as the mastery of the technique of reading leads to the production of sounds, and the correct sounds, when faced by print, so understanding, as mastery of the technique of language, leads, on the listener's part, to mental episodes, and the correct ones, when he hears certain well-formed sentences. There is no suggestion here that the understanding involves the private invoking of the formation rules of the language. One can agree with Wittgenstein that with complete mastery the operation with rules

becomes in some sense quite automatic, without abandoning the position that the end-product is an episode.

Nor does this conclusion conflict with his demonstration that, like reading, understanding is an achievement. Some achievement verbs (e.g., "win") are episodic, and, perhaps, most achievement verbs have both an episodic and a dispositional use, though in different contexts. This would explain why, even though "understand" has an episodic use, it seems inappropriate to ask "Did you understand continuously?"

V. MEANING, UNDERSTANDING, AND INTENTION

Let's turn now from episodic understanding by the listener of what the speaker means, to episodic meaning by the speaker of what the listener understands. Characteristically the speaker either intends that he should be understood in a certain way, or, more generally, it is at least intentional on his part that he should be so understood. Now an intentional performance is one which does not conflict with any of the performer's intentions, and the fact that it doesn't must be a necessary condition of its being performed. Accordingly, we can conveniently investigate meaning, and what is meant, in terms of the distinction between intending and what is intended, without ignoring the fact that all intentional behavior is not explicitly intended.

I shall now confine myself mainly to intentions to speak or write in a common language, without further reference to the listener's episodic sort of understanding which is a more ultimate aim of such behavior. This is partly to remain on common ground with Wittgenstein, but also to keep the discussion within manageable bounds.

Exposition—He uses roughly four distinguishable but interlocking types of argument against the claim that any intention can occur in a flash.

(a) How can we suppose that an intention can anticipate the whole usage of a language?⁴⁰ One may know, on sitting down to play a game of chess, what game one is about to play. But this is not based on the inductive ground that a certain sort of game is the usual consequence of a certain act of intending, nor on some "sort of

super-strong connection" between the two. More positively:

An intention is imbedded in its situation, in human customs and institutions. If the techniques of the game did not exist, I could not intend to play a game of chess. In so far as I intend the construction of a sentence in advance that is made possible by the fact that I can speak the language in question.⁴¹

(b) One may propose the rule "add 2" for the further expansion of the series "1, 4, 6, . . ." But what if on reaching 100 one's pupil continues with "104, 108, 112, . . ." One might, indeed, say to him "That is not what I meant," but this need not imply that one had at the time any such contingency in mind.⁴²

every interpretation together with what is interpreted hangs in the air, the former cannot give the latter any support. Interpretations by themselves do not determine meaning.⁴³

(c) If the contrast between voluntary and involuntary, e.g., my raising my arm and my arm merely rising, stems from an act of will peculiar to the former alone, then one can ask whether this act of will itself is voluntary or involuntary. Either answer would lead to difficulties.⁴⁴

(d) Consider, instead, the contrast in *grammar* within the prediction "I am going to take two powders now and in half an hour I shall be sick."⁴⁵ It suggests that the characteristic of voluntary action is simply the lack of surprise with which the agent views its completion: for the first conjunct is a characteristic expression of purpose, and, unlike the second, not based on inductive grounds, e.g., preceding thoughts or actions, far less upon some inner act of will (see (c) above).

Consider, again, why we find paradox in a statement like "I believe it will rain, but it won't."⁴⁶ Obviously "I believe it will rain" cannot function as a psychological report of some mental state. It must function as "It will rain" by itself, if it were used not merely to inform but also to manifest that the speaker is inclined to say it. This language game would be similar to that played by examinees in an examination. The functions of "I intend to . . ." and predictions like "I am going to take two powders now" are related in a

⁴⁰ *Ibid.*, I, §§ 188, 197, 337; cf. *Brown Book*, II, 5, p. 143.

⁴¹ *Investigations*, I, § 337.

⁴² *Ibid.*, I, § 187.

⁴³ *Ibid.*, I, § 198.

⁴⁴ *Ibid.*, I, §§ 613, *passim*.

⁴⁵ *Ibid.*, I, § 631; II, xi, p. 224.

⁴⁶ *Ibid.*, II, x, p. 191.

somewhat similar way. And what the prediction manifests is that a later report that he has taken two powders won't occasion surprise in the original speaker.

Wittgenstein extends similar analyses to statements that a speaker may make about his unfulfilled past intentions. If I say "When I was interrupted I was going to say . . .," the certainty with which I know what I was going to say doesn't require that I had thought it out, but only not said it, before being interrupted.⁴⁷ Nor need I have read what I *now* say I was going to say from some other process which I now remember took place then. Nor need I be interpreting the past situation and its antecedents.⁴⁸ To make these suggestions is "to look for an explanation where we ought to look at what happens as a proto-phenomenon . . . where we ought to have said: *This language-game is played.*"⁴⁹

He adds that the grammar of the expression "I was then going on to say . . ." is related to that of the expression "I could then have gone on" (cf. Section III). In the one I am remembering an intention, and in the other the having understood something, e.g., the formation rule of an arithmetical series.⁵⁰ He presumably means that reports about what one was going to do function as signals, exclamations, etc., in the same way as first person singular present or future indicative claims such as "Now I understand," "Now I can go on," and "I am going to do such-and-such."

Comment—We can concede in reply to (a) and (b) that no episodic mental state could have either the status of meaning or understanding something said unless the speaker had acquired the mastery of certain techniques. But, as I have argued in Section II, skills or know-hows are only realized in intentional actions (if we except the ability to understand episodically which perhaps should not be classed as a skill). Consequently we cannot concede the stronger conclusion toward which Wittgenstein seems to be reaching in the whole group of arguments, viz., that intentionality itself consists in some sort of mastery.

A further concession can be made to (b). "That's not what I meant: I meant . . ." may simply register surprise at the way a formation or syn-

tactical rule has been applied (see also Section III on "mean" and "fits"), without implying that the speaker had specifically provided against that particular application at some date in the past. But, as Wittgenstein remarks in (d), one's future voluntary actions are among the things which one predicts.

Finally, I accept (c) without reservation and cognate points in both (a) and (d). What makes an act voluntary is not some "inner" act of will; and the forecast of future action made in some expression of intention is not based inductively upon some causal connection between the agent's present state and the future action. But my reasons for this acceptance will become clearer in my rejection of the more positive analysis in (d).

To counteract any tendency, to construe the first person singular use of certain psychological verbs as that of reporting inner momentary states, Wittgenstein at several points adopts a principle which in its most explicit form runs as follows:

Do not ask yourself "how does it work with *me*?"

Ask "what do I know about someone else?"⁵¹ I believe one should take both questions in conjunction, but I don't wish to find fault so much with the principle as with the assumption that answers to the first question are more likely to postulate inner states. On the contrary, it is just the attempt to answer the second question in isolation from the first that leads most readily to the postulation of the mental as something inner. The inner-outer private-public dichotomy is one which comes most naturally to what I shall call "the alien," i.e., to the other person who has to consider my present states, or to myself when considering my own past or remotely future states. In my own immediate experience I can detect no box which in principle segregates me from something outside.

This can be seen most directly if we consider, first, the alien point of view toward an agent⁵² who may express his intention in the form "I am going to do *X*," from the alien point of view that if the agent is to be ultimately responsible for *X* coming about, then the following conditions must hold: (a) The circumstances of the action or standing conditions must be, at least roughly, as

⁴⁷ *Ibid.*, I, § 633.

⁴⁸ *Ibid.*, I, § 637.

⁴⁹ *Ibid.*, I, § 645.

⁵⁰ *Ibid.*, I, § 660; cf. I, §§ 180, 323.

⁵¹ *Ibid.*, II, xi, p. 206; cf. I, §§ 144, 147, 155.

⁵² For a fuller treatment see Rankin, *Choice and Chance* (Oxford, Blackwell, 1961), particularly Chapters VI-IX.

they appear to the agent when he makes his decision, (b) the circumstances must not in themselves be sufficient conditions of *X* coming about, and (c) in addition, for *X* to come about the agent must intend *X* to come about.

We may note here in passing that this account already departs radically from Wittgenstein's analysis of "I intend to do *X*" by analogy with "I believe it will rain," or of "I am going to do *X*" by analogy with "It will rain" as used to manifest something about the speaker. What the statement "It will rain" manifests is in no sense a condition of the fact that it will rain. On the other hand, what the expression "I am going to do *X*" manifests has something to do with *X* coming about.

But now the urgent question becomes "In what sense of 'additional,' or in what sense of 'condition,' can the agent's intending *X* be said to be an additional condition of *X* coming about?" It is here that it becomes so easy, if we keep the alien point of view exclusively in mind, to say that the intention is a causal condition like the various factors which constitute the relevant circumstances, except that it cannot be inspected by others, and hence must be something inner and occult. From the agent's point of view this is quite impossible.

(1) He cannot regard his intention as something additional to the relevant causal factors in the circumstances, because it is in consideration of what he takes these factors to be that he forms his intention. What, from the alien point of view, is an external relation between intention and circumstance must, from the agent's point of view, be internal, even where both are agreed as to what the relevant circumstances are.

(2) Again, the agent's intention is expressible in a declaration of the form "I am going to do *X*." Consequently, he cannot regard his intention as one of the causal factors which, in conjunction with those other causal factors in consideration of which he forms his intention, or in conjunction with any other, will bring about the occurrence of *X*. For *X* to occur the intention has to be formed, but the intention isn't formed until the agent is prepared to declare "I am going to do *X*."

I think at least two important conclusions follow from these arguments. (a) It is the alien alone who can make a distinction within the complex of action between subjective factors, like intentions, and objective factors, like the causal dispositions of physical objects: from the agent's

point of view the subjective-objective dichotomy just collapses. (b) Even from the standpoint from which the subjective-objective dichotomy is appropriate, one mustn't construe it as a dichotomy between inner and outer, occult and overt causal factors. Instead of ascribing conflicting ontologies to the agent and the alien, we must content ourselves with the obvious practical conflict which is implicit in our distinction between the two. Both, of course, are agents, but with respect to different prospective actions: and the conflict between their standpoints stems from the fact that each from the standpoint of the other is potentially passive in relation to the activity of others, and consequently as agents the active-passive organization of the world effected by the one must conflict with the active-passive organization effected by the other.

Here, however, I am in effect outlining a program⁵³ which I can only develop further now insofar as it affects Wittgenstein's principle paradox expressed by the question "How can I anticipate in advance either in some act of understanding or in an intention a whole usage?" This difficulty only arises if we interpret the mental-physical, subjective-objective dichotomy in terms of the inner-outer model. The only way in which we could regard an inner momentary state as anticipating an extended course of events in the future would be by virtue of some inductive generalization, and this, as Wittgenstein rightly points out, is not how we anticipate when we say "Now I can go on" or "I am going to do *X*."

If we abandon the inner-outer model, however, we see that we do quite definitely anticipate in a flash the whole of a future usage, but only in the entirely innocuous sense of "anticipate" involved by saying that a certain course of events involving, or possibly involving, myself is now in the future. The momentary state isn't some mysterious event wedged in between the past and the future, but simply the state, or cut, constituted by a certain course of events being now in my past, and another course, or set of possible courses, being now in my future.

To take the simplest sort of case, each past-future cut in my history differs at least in the mutual relationships of its components from the rest in the series. One out of a range of alternative actions, which in one cut constitutes what I could be about to do, is in a latter cut what I am

⁵³ I am currently working on a paper which brings the program nearer to completion.

actually going to do; and this gives any one of the actions it precludes the new status of that which I could have been about to do but am not about to do. We can now interpret the concept of intention as applicable *primarily* by each agent to other agents, and only in a derivative manner to himself at the executive moment of action. *Primarily*, if we simplify considerably, to speak of intention is to indicate the existence of other past-future cuts which are similar in general structure to, but controlled in content by, his own or others' in a way in which his own is controlled by none.

Of course, it is perfectly possible that the person who says "Now I can go on," or even "Now I am going to play a game of chess," may be making claims to which he is not entitled; and whether he is entitled to these claims can only be judged by others in terms of his past performance, and ultimately by what he does in fact do after making the claim. Indeed, what is intended may *sometimes* involve meeting certain social standards, and whether what the agent actually does will count as meeting these standards is dependent upon the actual usage, e.g., the day-to-day practice of the game of chess or a language, rather than upon some antecedent condition which guarantees success. But if we take these first person avowals as reports of the past-future cut at the time, rather than of some inner event intermediate between the two, then the fact that their correctness must depend upon what happens in the future, or upon what has just happened at some later date, is no longer a mystery.

At the same time the basis for Ryle's distinction between a way of operating or technique, on the one hand, and a usage, on the other, becomes more evident. It rests principally upon the distinction between an agent's past and future which I have just been vindicating. Roughly: a formula for behavior, social or otherwise (e.g., a formation or syntactical rule), describes a technique when the behavior in question is viewed prospectively; and the same formula may be taken as describing a usage if the behavior is viewed retrospectively. But more accurately: the formula describes a technique if the action in question is taken as belonging to the content, or possible content, of the future section of any past-future cut, irrespective of whether the cut itself belongs to the past, present, or future; and it describes a usage, actual or possible, if the action is taken as belonging to the content, or possible content, of the past section of any such past-future cut.

VI. CONCLUSION

Perhaps Wittgenstein's reputation is even yet in the somewhat turbulent and seminal phase in which criticism of his philosophy is more readily interpreted as denigration than as conscious tribute. It is worth declaring explicitly, therefore, what has, I think, been implicit in the organization of this paper, as well as in its more constructive conclusions: The deceptively simple question "How can one in a flash anticipate a whole usage?" is quite as profound as questions like "How can the one be many?" or "How can one event follow necessarily from another?" or "How can synthetic truths be *a priori*?" when these were mooted respectively by Plato, Hume, and Kant. On first inspection one might feel that the answer "One just does" has all the obviousness that a Wittgensteinian type of reminder should have, and Wittgenstein himself insists that in a sense it is obvious. But at the same time he has uncovered a natural tendency to interpret this reply in terms of models which, taken literally, are totally inappropriate, and this insight must have repercussions throughout the whole philosophy of meaning and mind.

I have argued, however, that he has replaced the misleading models by an anthropological one which has turned out in other respects to be equally misleading, even though it has the virtue of not conceiving understanding as some sort of source of correct usage. He has been misled by this model into over-simplifying the concept of understanding, or alternatively an over-simplification of the concept of understanding may have suggested the model to him.

By tracing an interconnection throughout the course of this paper between what Wittgenstein has to say on meaning, understanding, and intention, I have, of course, been trying to show something more than the way in which he misrepresents all three. In the final stages of my argument I have also tried to found a trilemma on this bond. Either meaning, understanding, and intention are to be interpreted in the Wittgensteinian manner, or they are to be interpreted as some sort of inner source of subsequent behavior, or the distinction between the mental and the physical must be regarded as secondary to that between the agent's and the alien standpoint in a way which disposes of the inner-outer dichotomy. Some of my destructive arguments dispose of the first horn; others, in conjunction with some

of Wittgenstein's destructive arguments, dispose of the second, and this would seem to leave the field to the third. Of course, this application of the method of elimination is a poor substitute for a detailed constructive proof that the distinction between the mental and the physical can be derived from the complementarity between the agent's and the alien standpoint in some way which does not surreptitiously reinstate the inner-outer dichotomy, but what is no substitute can yet serve as a useful preliminary.

Finally, it may be⁵⁴ that no theme of Wittgenstein's ever achieves, or was intended by him

to achieve, complete consummation as a thesis or set of theses. But whatever truth there is in this should not lead one to overlook the delicate adjustment whereby each part within the total assembly would seem to support some of the weight of the rest, if only on the strength of a mutual relevance distributed in criss-cross fashion⁵⁵ fairly evenly throughout—and it is with these sorts of interconnection that I have concerned myself, both in the more negative and the more positive parts of my program, rather than with the weight which might be placed on any terminal point.

Monash University

⁵⁴ P. K. Feyerabend, "Wittgenstein's Philosophical Investigations," *Philosophical Review*, vol. 64 (1955), pp. 449-483; Judith Jarvis Thomson, "Private Languages," *American Philosophical Quarterly*, vol. 1 (1964), p. 36.

⁵⁵ *Investigations*, p. ix.

II. AGENT CAUSALITY

JOHN W. YOLTON

RECENT attention to the concepts of intending, willing, and deciding, together with concern for the role of the emotions and knowing in practical judgment, has pointed the way to a philosophy of action freed from some of the excesses of earlier approaches to these concepts. What we do not have as yet is a philosophy of the person adequate for action and moral responsibility. A related lack is the absence of any analysis of cause adequate for knowing and doing. Both the agent and the agency of action need a bolder analysis than they have yet received. In this paper I shall try to suggest a viable notion of agent causality by (a) stressing the contributions toward understanding the notion of practical knowing made by analytical and phenomenological philosophers (Section I); (b) exposing and questioning the substitution of logical for causal relations in action, found in Hume and in some recent writers (Sections II–III); (c) developing the notion of mental causes, making use in particular of an important chapter in Johnson's *Logic* (Section IV); and (d) showing how agency depends upon cause as the *initiator* of consequences (Section V). For this last point, I think it helpful to remind ourselves of Kant's treatment of agent causality.

Throughout this paper I have allowed the points I want to make under each section to emerge from my analysis of other writers; the truths and the mistakes of the writers I discuss are made to speak for me. Any reader who wants to know my main conclusions may consult the final paragraph of each section, where I have stated my own conclusions.

I. PRACTICAL KNOWING

Miss Anscombe has rightly called attention to the inseparability of volition and sensible discrimination.¹ Cognition (certainly at its lower levels) is controlled by the needs and interests of the agent. The sign of wanting is "*trying to get*" (A, p. 68). By this phrase, Miss Anscombe does not mean any sort of disposition to respond in

specific ways; she recognizes that in saying, of an animal or of a person, that he is trying to get something, we are describing the movement "in terms that reach beyond what the animal" or person is now doing (A, p. 68). One of the ingredients in wanting is "movement towards a thing and knowledge (or at least opinion) that the thing is there." Moreover, wanting eventuates in action, might even be said to be a cause of action, if we can elaborate a notion of cause freed of Hume's conjunctions and consistent with Kant's talk of free causality. Since wanting is one side of a coin, the other side of which is a kind of knowing, an understanding of this knowing may help in the construction of the concept of agent causality.

Much mystery has surrounded Miss Anscombe's notion of practical knowing. She herself characterizes this knowing as "knowledge without observation," referring initially to a bodily phenomenon, e.g., our knowing the position of our own limbs. Her main concern—though this fact has been overlooked in most discussions of this part of her book—has been with the knowing that goes with willing and intending. She begins her discussion of knowledge with observational knowing. She wishes to show that when I know what I am doing, this knowing is not observational: does not rest upon some special sensation which I inspect. She rightly charges Locke and Hume with fostering the inspection theory of the knowledge of our internal states (A, p. 76). The outcome of the inspection theory, when joined with the view of knowing as the observation of some thing or event, is a paramechanical view of self-knowledge. We are driven to "look for" special sensations, efforts of will, etc. (Cf. A, p. 57).

Miss Anscombe wants us to recognize a non-observational mode of knowing: practical knowing. We could equally well call this mode "intentional knowing"; the virtue of her use of Aristotle's and St. Thomas' label of "practical knowing" is that it links this knowing immediately to doing, or to action. It is most important to understand that practical knowledge is not doing itself, nor

¹ *Intention* (Oxford, Blackwell, 1957), p. 68. Hereafter, this work will be referred to by the capital letter A.

any disposition to do. Miss Anscombe emphasizes that in ascribing practical knowledge (it is the other side of the coin, wanting, that she stresses) to any animal or person we have gone beyond what is observable. She is committed, I think, to treating knowing as some sort of mental event, but her use of the concept of practical knowing is not directed toward the knowing or the wanting as mental processes. All she wants to establish is that an agent knows what he does, not because he observes himself doing what it is he does, but simply because *he* does it. The form of description of some action is acceptable by the agent because he knows what he has done. "What is necessarily the rare exception is for a man's performance in its more immediate descriptions not to be what he supposes. Further, it is the agent's knowledge of what he is doing that gives the description under which what is going on as the execution of an intention" (A, p. 86). To give an account of what *I am doing*, as opposed to giving an account of what my body or some part of it is doing or undergoing, I do not have to resort to observations, though observations may play an important role in enabling me to execute my intentions (A, p. 88). As Vesey puts the point, "a person can know that his hand moved, when he moved it, solely by virtue of the fact that *he* moved it."²

In short, in instances of my moving my body I do not "first have to learn of the association of one thing, some quality of a mental event, with another, some disturbance or movement of some part of my body. . . . My awareness of myself as embodied is not *mediated*."³ The mystery around Miss Anscombe's talk of knowing the position of my limbs without observation is dispelled once we pick up Vesey's cue about the embodied self. Knowing what I am doing in cases of actions proper—signing a contract, repaying a debt, telling the truth—is akin to the knowledge of one's own body; it is akin to and perhaps even derivative from such knowledge, though I am not sure it is quite the same. Wanting and intending both presuppose and are forms of knowing. To sense is to make sensible discriminations. Similarly, I cannot want without knowing what I want. It is not that the wanting and the knowing are separate or even distinguishable. Similarly, to intend, or to

have an intention, is not something we can do or have unknowingly. The knowing in these cases is a species of knowledge without observation. The other species of this same sort of knowledge cited by Miss Anscombe is the knowledge we have of the conditions of our own body. Vesey's cue for the latter species of knowledge is phenomenology. Vesey uses Sartre's formulation of the "body for us" notion (as opposed to the body-in-itself, as an object for observation).⁴ Merleau-Ponty's analysis is even more effective and more related to our topic.

Vesey's example is that pain is not located by observation. Not only is the location of pain not a matter of observation, but the body itself, in so far as it is *my* body, is not an object to be observed. Observation presupposes that I have a body; my body is my way of having a world. "In other words, I observe external objects with my body; I handle them, examine them, walk around them, but my body itself is a thing which I do not observe: in order to be able to do so, I should need the use of a second body which itself would be unobservable."⁵ A shorter way of saying the same thing is to say that in so far as my body "sees or touches the world, my body can therefore be neither seen nor touched" (M-P, p. 92). This fact about the body defining for us a world, and not itself being an object in the world, is part of what Merleau-Ponty calls "being-in-the-world." The other component in our being-in-the-world is the cognitive one. The fact that my body is not an object of observation does not preclude my knowing about my body; I simply do not know it by observation. The kind of knowing operating here is like (though not the same as) the "taking account of" exemplified by animals in leaping for prey, or in my being able to walk up a set of stairs without thinking what to do next. This knowing is forcefully illustrated in Merleau-Ponty's analysis of the phantom-limb phenomenon. This phenomenon is also helpful in revealing the role of the agent in action.

It is helpful to remind ourselves that seeing is not just a matter of the physical organs of sight operating properly, though they must be in working order before I can see. Seeing is a psychological, or better, a psychophysical operation. There is no simple correspondence between the

² G. N. A. Vesey, "Knowledge Without Observation," *The Philosophical Review*, vol. 72 (1963), p. 209.

³ *Ibid.*, pp. 211-212.

⁴ "The Location of Bodily Sensations," *Mind*, vol. 70 (1961), p. 32.

⁵ M. Merleau-Ponty, *Phenomenology of Perception* (London, Routledge and Kegan Paul, 1962), p. 91. Referred to hereafter as M-P.

proper stimulus and what I see. Seeing organizes and structures stimuli. Different neurophysiological factors can respond to the same stimuli in the same way, if old pathways have been destroyed, for example. Most, if not all, actions are psychophysical. It is misleading to "look for" the psychological component. Merleau-Ponty has said it is wrong to "look for" the physical component of action when the action is my own. The conflation of knowing into doing is equally misleading if it is taken as denying the two components. Vesey's suggested formula,⁶ "So far as he, but not necessarily his arm, was concerned, he raised his arm" is wrong; but the reasons why this is wrong need more careful expression. Another move Vesey tries is that what *I do* in raising my arm (as opposed to what happens to my arm) is indicated by "I raised my phantom arm."⁷ The phantom-limb phenomenon does shed light on the psychological component of action, though not in the way Vesey suggests by this formula.

In understanding this phenomenon we must recognize that "a collection of cerebral symptoms could not represent the relationships in consciousness which enter into the phenomenon" (M-P, p. 76), for psychic determinants initiate and maintain the phenomenon. At the same time, "no psychological explanation can overlook the fact that the severance of the nerves to the brain abolishes the phantom limb" (M-P, p. 77). We are familiar, through the teachings of phenomenologists, with the notion of the *Lebenswelt*; there are reasons for speaking of a quasi-knowledge which characterizes an animal's relation to its world.⁸ What such quasi-knowledge requires is a bodily recognition of the stimuli relevant to that world. Each stimulus takes on a significance for the animal only as part of the "global situation" which is its world (M-P, p. 79). Psychic and physiological factors work together in human actions. The patient suffering phantom-limb phe-

nomenon furnishes an interesting instance of such a combination. He is not ignorant of the paralyzed or absent limb: "He can evade his deficiency only because he knows where he risks encountering it, just as the subject, in psychoanalysis, knows what he does not want to face, otherwise he would not be able to avoid it so successfully" (M-P, p. 80). It is in the same way that the anosognosic "leaves his paralysed arm out of account in order not to have to feel his handicap, but this means he has a preconscious knowledge of it" (M-P, p. 81). In short, "The man with one leg feels the missing limb in the same way as I feel keenly the existence of a friend who is, nevertheless, not before my eyes; he has not lost it because he continues to allow for it, just as Proust can recognize the death of his grandmother, yet without losing her, as long as he can keep her on the horizon of his life. The phantom arm is not a representation of the arm, but the ambivalent presence of an arm" (M-P, p. 81).

Similarly for normal action: in order to walk, I do not need a clear and articulate perception of my body; it need only be "at my disposal." Walking involves a kind of knowing, the same kind of knowing Miss Anscombe talks about in knowing the position of one's limbs. It is in this way also that wanting is a kind of knowing. The difference between the knowing of wanting and the knowing of walking is that the former is a much more conscious knowing. To do what I intend or want, I must, of course, have my body at my disposal, since it is in the field of the body that my intentions are fulfilled and expressed. What I know in action is not, primarily, what my bodily movements are; what I know then is what *I am doing*, repaying a debt, not just extending my hand to you.⁹ The kind of knowledge I have of my body is also the kind of knowledge I have of my self, of my intending, wantings, etc. Just as I will not understand "the functions of the living body except by

⁶ "Volition," *Philosophy*, vol. 36 (1961), p. 363.

⁷ *Ibid.*, p. 353.

⁸ For some further discussion of this point (together with some comments on phenomenological analysis), see my "The Form and Development of Experience," *Acta Psychologica*, vol. 21 (1963), pp. 357-363.

⁹ I have said more about what actions are in "My Hand Goes Out To You," forthcoming in *Philosophy*. One of the more interesting features of recent analyses of action is the way in which the English channel has been bridged. A paper of fundamental importance is "Observation and the Will," by Brian O'Shaughnessy, *The Journal of Philosophy*, vol. 60 (1963), pp. 367-393. In that paper, O'Shaughnessy makes use of many phenomenological concepts, without any suggestion that he has consciously borrowed them from continental philosophers. He makes effective use of the distinction between *the world* and *my world*, locating action in the latter. O'Shaughnessy's paper should be read in conjunction with Merleau-Ponty's book on perception. See also P. Winch's "Understanding a Primitive Society," *American Philosophical Quarterly*, vol. 1 (1964), p. 323: "The life of a man is a man's life and the life of a woman is a woman's life: the masculinity or the femininity are not just components in the life, they are its *modes*. Adapting Wittgenstein's remark about death, I might say that my masculinity is not an experience in the world, but my way of experiencing the world."

enacting it myself" (M-P, p. 75), so I cannot treat the psychic components of action within the categories of objective nature. The category of the "in itself" must be replaced by that of the "for itself." The classical psychologist's attempt to write the history of consciousness in objective terms must be replaced by the recognition that "This history of the psyche which he was elaborating in adopting the objective attitude was one whose outcome he already possessed within himself, or rather he was, in his existence, its contracted outcome and latent memory" (M-P, p. 96).

Practical knowing, then, is the knowing needed before I can be an agent in action. Observation is something *I* do, but the knowledge I acquire via observation is a product of my doing directed upon the world. The knowledge I have of my intentions and of my doings is something I acquire, it is something I have because it is I who am acting and my actions are what I know. I cannot be an agent and fail to have such knowledge.

II. LOGICAL RELATIONS IN ACTION

Most recent analyses of action exhibit a curious reluctance on the part of their authors to employ the notion of cause. The role of mental events in actions—of intending, understanding, etc.—has only been granted in an oblique way. Miss Anscombe defined intentional actions in terms of language (A, p. 85), though she did recognize the need for a philosophical psychology. What she had in mind in referring to her analysis as linguistic was the role of description and of forms of description, in our identification of actions. For her, the term "intentional" refers not to some interior occurrence (certainly not primarily), nor to observable behavior, but to a form of description. The form of description is a reflection of the rules and customs of a way of life.¹⁰ MacIntyre has made the same point more emphatically. One of the basic conditions for any event's being an action is that the event fall "under some description which is socially recognizable as the descrip-

tion of an action."¹¹ From this it follows that "an agent can do only what he can describe."¹² MacIntyre uses this notion as a way of showing the importance, in sociology, of understanding the ideologies of any given society: "To identify the limits of social action in a given period is to identify the stock of descriptions current in that age."¹³ The "stock of descriptions" is the ideology of the society.

The description of some bit of behavior in action-locutions recognizable by the agent is one way to pinpoint just what has been done. But if we stop our analysis of action with the description, "he pushed upwards very hard," "he repaid his debt," we stop short of a theory of action. A theory of action should be not only descriptive in this sense, but it ought to tell us in more detail what has been done, how it was carried off, and just how it differs from physical movement. We are not told enough when it is said the difference between a certain motion of my hand being that motion and its being the signing of a contract is just that it is the signing of a contract. What this tells us is that the request for physical accounts, for muscle talk, is out of order. It does not tell us how the signing, as signing, was possible. The description of some action does tell us a bit more. It tells us that what was done was rule-following. The descriptions of actions are descriptions of events in which rules and the following of rules are operative. Those who talk of rule-following behavior tend to suppress the fact that to follow a rule, whether consciously and explicitly or not, requires certain sorts of psychological events.¹⁴ Even Winch, in an otherwise perceptive study, tends to settle for the magic of the phrase "rule-following," not making much of his own recognition that to follow a rule presupposes that I am *aware* of what I am doing, that I *understand* and have some *conception* of the significance of my action.¹⁵ The *description* of an action is in third-person terms; *my doing* involves first-person operations.

Besides analyzing action in terms of the appro-

¹⁰ For a fuller discussion of Miss Anscombe's analysis of action, together with some diagnosis of the contemporary approach to these problems, see my "Act and Circumstance," *The Journal of Philosophy*, vol. 59 (1962), pp. 337-350.

¹¹ A. MacIntyre, "A Mistake About Causality in Social Science," *Philosophy, Politics and Society* (2nd Series) ed. by Laslett and Runciman (Oxford, Blackwell, 1962), p. 58.

¹² MacIntyre, *ibid.*, p. 59. Cf., "Roughly speaking, a man intends to do what he does" (A, p. 45).

¹³ *Ibid.*, p. 60.

¹⁴ Some earlier attempts to discredit the mentalist features of action, those of H. L. A. Hart and A. I. Melden were criticized by me in "Ascription, Description, and Action Sentences," *Ethics*, vol. 67 (1957), pp. 307-310; also in "Act and Circumstance," *op. cit.*

¹⁵ P. Winch, *The Idea of a Social Science* (London, Routledge and Kegan Paul, 1958), p. 51.

priate description and rules, there is another way in which talk of causes is avoided in talking about actions: logical relations are substituted for causal ones. Because he is concerned to point out that the standard Humean conditions for causal ascription are wanting in human action—we do not have constant conjunctions between beliefs or intentions and specific sorts of actions—MacIntyre characterizes the relation between belief and action as logical. I think what he means by this characterization is that there is an internal, i.e., conceptual, relation between belief and action.¹⁶ But his elucidation of what he means by “logical” and “internal” is misleading. What I believe he has in mind is the sort of point Winch has made, that “an event’s character as an act of obedience is *intrinsic* to it in a way which is not true of an event’s character as a clap of thunder.”¹⁷ The point Winch is making is similar to the one about action being circumscribed by the available descriptions. To say of some agent that he obeyed a command, assumes that there are, in his society, commands, and that he understands the notion of a command. “An act of obedience itself contains, as an essential element, a recognition of what went before as an order.”¹⁸ This sort of “containing,” by the act of obedience, of a recognition on the agent’s part of the prior command is similar to the “containment” of predicate in subject in analytic propositions. Moreover, we might extend the similarity to implication. Winch even advances the interesting suggestion that logical relations between propositions are derivative from social relations between men.¹⁹

Referring to Aristotle’s doctrine of the practical syllogism, MacIntyre applies logical talk to the way actions “follow from” the premisses of the practical syllogism. He claims that the relation of conclusion to premiss in theoretical syllogisms is the *same* as the relation between the action and its premisses in the practical syllogism. To accept the premisses of a practical syllogism but to refuse to act is exactly the same as accepting the premisses of a theoretical syllogism while denying the conclusion: “we literally do not know in the one case what he is saying and in the other case what he is doing.”²⁰ MacIntyre’s brief use of the practical

syllogism falls prey to the very danger Miss Anscombe underlines in her *Intention* (33–45): he fails to call attention to the importance of the motive to action which supports the first premiss. What is described in that premiss “must be wanted in order for the reasoning to lead to any action” (A, p. 65). Later on, Anscombe refers to a “desirability characterization” to which the agent must respond. MacIntyre does speak of the believer “accepting” the premisses, but he is unilluminating as to the nature of that acceptance. Miss Anscombe’s point is that the acceptance must be one of wanting. An intellectual acceptance (as in theoretical syllogisms we may accept the truth value of the premisses) is simply not sufficient to induce action. Acceptance of the truth of the premisses is not the same as acceptance of the goodness of what the premisses talk about (A, pp. 75–76). To call the conceptual link, e.g., between a command and an act of obedience, “logical,” and then to compare this relation to the way premisses and conclusions are related in ordinary syllogisms, not only misleads but fosters the illusion that we can get along without “cause” in our analysis of action.

Kenny’s recent use of the notion of non-contingent (i.e., necessary) relations between emotions and their objects is the most extended effort to replace causal by logical (or conceptual) relations.²¹ Kenny applies the contingent-necessary terminology to certain historical figures as well as to the question itself of emotions, their objects and causes. One of the main issues Kenny wants to attack is the claim that an emotion can be recognized without its objects and circumstances. We must know the cause of some response before we can know the response to be an emotional one, or of some other kind (p. 44). According to Kenny, Descartes says that an emotion is only contingently related to its manifestation in behavior and to its cause (p. 12). We can be certain about this emotion while doubting the existence of its bodily manifestation. The certainty of the inner state makes its relation to outward behavior only accidental and contingent. This applies similarly for Hume. “It is because our minds happen to be made as they are that the object of

¹⁶ MacIntyre, *op. cit.*, p. 52.

¹⁷ Winch, *op. cit.*, p. 125.

¹⁸ *Ibid.*

¹⁹ *Ibid.*, p. 126.

²⁰ MacIntyre, *op. cit.*, p. 53.

²¹ A. Kenny, *Action, Emotion and Will* (London, Routledge and Kegan Paul, 1963).

pride is self, not because of anything involved in the concept of pride . . ." (p. 24). Contingency leads to inductive knowledge, not certain knowledge. Kenny wants certainty but not at the price of divorcing emotion from its manifestation.

"A feeling of anxiety is non-contingently related to the alarming circumstances which give rise to it; the relation between dyspepsia and its cause is purely contingent" (p. 84). Later, in dealing with pleasure, Kenny comes to a similar conclusion. If I am performing two independent actions at the same time, I am never in doubt about which of them I am enjoying (p. 131). Kenny says that doubt would be possible here "if pleasure were a sensation concomitant on and merely contingently connected with action." He wants to attack the notion of specific sensations which *are* the pleasure in each case, for if we were to say "what makes *any* event or action pleasant is its being accompanied by one or another of a set of specific sensations" (rather than simply that some sensations may be classified as pleasant), we would thereby be making the "connection between the particular sensations and the objects which produce them" contingent and hence the connection would have to be learned by experience (p. 133). The fact that the pleasure of drinking does not occur while we are eating "nor the pleasure of climbing the Matterhorn while one was toasting crumpets" are non-contingent facts or relations. The question is whether these non-contingent relations are facts about the emotions or facts about our talk and thought about the emotions. Is Kenny arguing a thesis about the nature of the emotions, or one about our talk of the emotions, or perhaps one about the conditions for our ascribing emotions to some person?

Kenny may deny the meaningfulness of the distinction between facts about emotions and facts about our talk of the emotions. Like most conceptual analysts, he moves frequently and easily from object to concept, from "it is no part of the concept x that ϕ " to "it is no part of the *nature* of x that ϕ ." The peculiarity of talking of necessary relations in nature may be clouded over by this move from nature to concept. Kenny's chapter on "Actions and Relations" may be an attempt to get clear about this difference. He there says that "Leonardo da Vinci painted the Mona Lisa" is a necessary proposition because "the Mona Lisa is essentially a painting by Leonardo da Vinci." That is, this is a necessary proposition if taken as a sentence about the Mona Lisa. If taken as being

about Leonardo, this same proposition is contingent since there is nothing essential about painting this picture for Leonardo. This proposition, interpreted as being about the Mona Lisa, does seem to express something essential to that painting, but even here it is not clear whether this is a fact about the painting or a fact about our talk about that painting. It is true that the Mona Lisa is a painting by Leonardo, but this fact hardly makes the proposition necessary that "Leonardo painted the Mona Lisa." Conceivably, there could be two such paintings: two paintings with the same content, the same color, etc. We could talk about both paintings, even compare them. We might discover that they were painted by two different artists, Leonardo and a clever forger. Only if we mean by the phrase, "the Mona Lisa," the painting by Leonardo, could the proposition above be a necessary one: *that* painting is Leonardo's.

Could we say that a relation or fact is necessary in this sense, in all cases Kenny uses, prior to the Leonardo example? One could construct four parallel propositions, patterned after the Mona Lisa one, using examples Kenny elsewhere uses. (1) "When I have inner state S_1 , I utter the word 'fear'" (p. 69). (2) "I feel anxiety under circumstances C_1 " (p. 84). (3) "Pleasure accompanies actions of this and this sort" (p. 131). (4) "The pleasure of drinking occurs when I am drinking" (p. 138). In each of these four cases, I do not think there is anything necessary about the relation between the feeling of anxiety or the pleasure or the inner state and the objects and circumstances mentioned in the proposition. That is, the application of the notion of a necessary connection between feeling and circumstance is inappropriate for the very reason cited by Hume: the contrary of any matter of fact is always possible. But, and I take this to be Kenny's point, there may indeed be something more than contingent about *our talk* of these feelings and their circumstances. Proposition (2), for example, is factually true; but it also has a necessity to it, in the sense that, if I say I feel anxiety, but the conditions are such that they are incompatible with anxiety-feelings, we either would not understand my remark, or we would have to view it in a psychoanalytic context. The range of circumstances and causes compatible with anxiety-feelings need not be precisely or particularly spelled out, but the type must be; otherwise we do not have the notion of anxiety. "Compatible" is linguistically defined. Similarly for proposition (5), "I can see only with the eyes."

This proposition is factually correct, but the conceptual truth here is such that, if someone told us there was a race of men (or even one individual) who saw with their ears, we would not understand what this meant. (It might mean that these men made predictions about the visual properties of objects, predictions which we could check. But then the claim is one about predictive powers, not about the experience of seeing. What we mean by "see" involves using the eyes.)

Kenny borrows the notion of a *formal object* as a way of making more forceful his claim about necessary relations between emotions and their objects. A formal object of any act of ϕ ing is "the object under that description which *must* apply to it if it is to be possible to ϕ it" (p. 189). For example, only what is edible can be eaten: "thing which is edible" is the formal object of eating. A more significant example is "only what is dirty can be cleaned," "only what is wet can be dried," or "I can steal only other people's property." The relation between an emotion and its formal object (and any act and its formal object?) is a logical relation (p. 191). The material object of any act is the specific object involved in that act. "Thing which is seen" is the material object of seeing; "this table" is the material object of this specific act of seeing. But the formal object seems to be not an object proper, not something in the world (as it were), but only a name for a form of description. "To assign a formal object to an action is to place restrictions on what may occur as the direct object of a verb describing the action" (p. 189).

Kenny's thesis about formal objects and necessary connections turns out to be a thesis about the conditions for discourse, for ascribing and talking about emotions. It is a claim about the language of emotions. An emotion which was private, only contingently related to some object or behavior, could not be understood. If he could be convinced that the notion of private mental events could be given a meaning, that we could learn the meaning of mental words even though the relations to object, circumstance, and behavior were contingent, would Kenny change his mind about emotions having a formal object located in circumstances? If he does not reach his view of the relation between emotions and objects from a belief that private meanings for emotion words is the only alternative (and that such meanings

could never be learned), how does he arrive at his thesis? If his thesis is only about formal objects, it is not saying much that is controversial. He seems to want to say that the only meaning we can understand for emotions is one in terms of public objects and circumstances. As in the case of most recent writers on action, a theory of meaning controls the analysis, turning it into one about language and action-ascription.²¹ Logical relations in language have replaced causal relations in action.

III. CAUSE AS A LOGICAL PROPERTY

The state of confusion in which the notion of cause has been left (A, p. 10) is a consequence of the analysis of cause as a uniform and ordered sequence, the analysis initiated by Hume and refined by Kant. This analysis took the "force" out of "cause." It also made our knowledge of cause the result of repeated observations, of induction. Agent causality requires force and influence, as well as non-inductive knowledge.

Hume's analysis of "cause" is bound in with his deductive ideal of knowledge.

There is no object which implies the existence of any other, if we consider these objects in themselves, and never look beyond the ideas which we form of them. Such an inference would amount to knowledge, and would imply the absolute contradiction and impossibility of conceiving anything different.²²

Because such deductive inferences from one object to another are impossible for man, the only resort is to causal, i.e., customary, inferences. If reason was the basis for causal inferences it would have to proceed on the principle "that instances, of which we have had no experience, must resemble those of which we have had experience and that the course of nature continues always uniformly the same" (p. 91).

Again the appeal to the possibility of conceiving the denial of this principle is used as one of the ways of showing the lack of demonstrative proof for that principle. What is logically possible may be factually so; a genuine knowledge of nature would coincide with what is logically necessary. Hume's analysis of causation reflects this logical concern with knowledge. It is not *necessary* "that everything whose existence has a beginning, should also have a cause," because we *can* conceive without

²¹ I have argued this charge in "Act and Circumstance" (*op. cit.*).

²² Hume's *Treatise* (Everyman Edition), p. 91. Unless otherwise noted, references to Hume are to this work and this edition.

contradiction the beginning of some event without conceiving of a cause (p. 81). Since we can find no reason why this particular event *must be* the cause of that particular event, we cannot say "that such particular causes must necessarily have such particular effects" (p. 81). So captured was Hume by the deductive model of knowledge that he was led to claim near synonymy for "necessity" and the usual causal terms, "efficacy," "agency," "power," "force," "energy," "connection," "productive quality" (p. 155). The denial of the conclusion "that there must somewhere be a power capable of producing" the several "new productions" in nature which we observe, is primarily that reason "can never make us conclude that a cause or productive quality is absolutely requisite to every beginning of existence" (p. 155).

Even in the *Enquiry*, this deductive model is present. "From the first appearance of an object, we never can conjecture what effect will result from it. But were the power or energy of any cause discoverable by the mind, we could foresee the effect, even without experience; and might, at first, pronounce with certainty concerning it, by mere dint of thought and reasoning."²⁴ How else to account for such a striking confusion between the idea of cause and an *a priori* knowledge of the effects of causes, given the causes, save to remark that Hume has assumed that cause is a logical property from which deductions could be made. A number of questions are involved here. One question is: "What would be the epistemic consequences were we to discover causes in nature?" The expected answer to this question would be that we might be able to make predictions, but hardly *a priori* or *deductive* predictions. Another question is: "In order to have an idea of power, must I discover instances, and recognize them as instances, in nature?" Hume's answer seems to have been "yes," but in so replying he has surely overlooked the fact that many thoughts about the nature of the world are not thoughts which have any *known* correlate in the world; he assumes that all ideas are descriptive ideas, and known to be so. Another question is: "If we have an idea of cause or power, how does it arise"? To

this question, Locke has supplied an answer in terms of the sorts of experiences which elicit the reasonings which lead us to conclude that events have causes.²⁵ Locke's derivation of the idea of power was not a derivation of the idea of necessity. The notion of some object producing or determining or bringing into being some other object or event is not the same as the notion of logical necessity. Hume seems to have confused logical implication with causal influence. Either Hume is denying that *in fact* we have an impression of force, production, influence, or he is denying that objects and events are related by logical connectives.

Hamlyn interprets the first alternative as Hume's way of pointing out that "being the cause of something is not a perceptual property of an object in the same way that redness is."²⁶ Hamlyn claims that this is a conceptual point, and hence immune from experimental verification or rejection. Earlier, Hamlyn had distinguished between the ordinary sense of "impression" and Hume's "more technical sense," but it is difficult to see what the difference is for Hamlyn. He claims that Michotte, in the latter's work on causation,²⁷ is working with the ordinary notion, hence that Michotte's studies do not refute Hume. Hamlyn further says that Hume's thesis about our notion of "cause" is "compatible with the thesis that in some sense of the word 'impression' we do have an impression of causality" (p. 77). Hamlyn is, I believe, correct in saying Hume has so defined "impression" as to make it a technical term in his language: this is the only sense in which Hume's analysis can be said to be conceptual.²⁸ If "impression" is used as "perceptual quality," and if that is the technical meaning of the term in Hume, I do not think we could find anyone who would have asserted what Hume denies in his analysis of "cause." At least, I have been unable to find, among Hume's contemporaries or immediate predecessors, any who claimed or assumed that "cause" or "power" designated a property of objects in the way "red" does.

I find it difficult to understand Hamlyn's quick rejection of Michotte's investigations. The place

²⁴ Selby-Bigge, 2nd ed. (London, Oxford University Press, 1902; reprinted 1963), p. 148; cf. p. 150.

²⁵ See his *Essay*, Book II, chap. xxi, § 1.

²⁶ D. W. Hamlyn, *The Psychology of Perception* (London, Routledge and Kegan Paul, 1957), p. 79.

²⁷ A. Michotte, *La Perception de la Causalité* (Louvain, Institut Supérieur de Philosophie, 1946). (English translation by Miles and Miles; London, Methuen, 1962.)

²⁸ Just because Hume's analysis is conceptual (not only in what he has to say about "cause"), his analysis of experience—even the concept of experience itself—has a strong *a priori* flavor. See my "The Concept of Experience in Locke and Hume," *Journal of the History of Philosophy*, vol. 1 (1963), pp. 53-71.

335077

to start in any analysis of causality would be, I would think, with our ordinary notion of cause. Michotte has supplied us with some effective analysis of that ordinary notion. What Michotte's experiments remind us of—what they in fact depend upon—is our own ordinary everyday perception of causal situations. The notion of "cause" has a use in our language; it is learned in our society, and it has roughly the sense of "power" or "production." Michotte's experiments by no means invoke, as Hamlyn charges, the "assumption that . . . we may isolate the 'pure experience'" (p. 78). Michotte's subjects could not have responded as they did under experimental conditions if they had not already been making causal ascriptions before coming to the laboratory. I would presume that this ordinary notion of cause was as much a part of the responses of men in Hume's day as in our own. It was this notion which Hume partially accounted for in relating it to imagination, feeling, and expectation.

Hume's account of our notion of cause was not only partial, it was one-sided as well. That is, Hume did not have much to say about the object side of cause; his principles led him to restrict his analysis to the idea of cause, to the psychology of cause-ascription.²⁹ What it means for us to say "*X* causes *Y*" is that we have a feeling of constraint and anticipation when confronted with future *X*-like events. The ordinary notion of cause includes some thoughts about the causal process as well, some thought of *X* producing *Y*, of *X* bringing it about that *Y*, or *X* initiating the action of *Y*. Philosophers have trained themselves away from such notions of cause; science has reinforced the substitution of correlation. But the notion of cause as correlation, as ordered sequence, is clearly inadequate when we seek to formulate our thinking about action. Neither is it a notion adequate to our ordinary experience of pushing, exerting, exploding, etc. To say "*X* causes *Y*" usually involves the thought of energy expended, of chemical or electrical processes, of motion and impact. There may be no need to retain these notions of cause in our sophisticated philosophy of nature or of science. We need them in our philosophy of man. Yet, paramechanics threatens

if we try to apply these pushings and impacts, even these electrical and chemical changes, to thought and action.

IV. MENTAL CAUSES

There are three sorts of mental causes. (1) There are those mental causes which are physical events affecting a sentient creature. (2) There are those which require understanding by the person affected, as I must understand the meaning of the joke before it can make me laugh. (3) There are those mental causes which are the causes of action which may be improperly called "mental": they might better be called agent or actor causes. The third sort of mental cause is Kant's *free* cause. I shall be arguing that this third sort is bound up with the person and that we cannot analyze this kind of cause without an analysis of the person. To be a person is, in part, to be the agent of action. To be the agent of action is to be, in part, the cause of those actions.

Agent causality is, however, a *type* of mental cause. It also presupposes some of the other kinds of mental causes since it presupposes awareness. Can we say any more about mental causes in sense (1)? Teichmann says³⁰ that the object seen could not be the cause of the seeing. Such a remark is true if "thing seen" be taken in the awareness sense, though such a remark needs qualification. "Thing seen" is the end result of the complex, physiology plus awareness. The process in full is as follows: physical event or object (analyzed as light or sound waves or color waves, etc.), neurological events, eye, awareness, visual sensations, object seen. The latter is clearly different from the physical object. The physical object may be said to be the cause of the neurological events, but only under the condition of awareness or consciousness are the events fully possible. The very fact that consciousness and attention are correlated with different brain waves through inattention or sleep suggests that consciousness is a necessary condition for those neurological events, necessary for visual sensations. What we seem to have here is a close intermingling of physical and mental events in the production of a visual experience.

²⁹ Harré has pointed out that Hume confuses the *criteria* for asserting a causal relation with the *meaning* of that relation. The Humean tradition, Harré further remarks, reduces "the causal relationship from an internal connection between cause and effect to a summary expression of an external relation between events." Harré is correct in saying, "What we mean by '*X* causes *Y*' is that *X* generates (or produces) *Y*, not that *X* is followed by *Y*." See his "Concepts and Criteria," *Mind*, vol. 73 (1964), pp. 354-360.

³⁰ J. Teichmann, "Mental Cause and Effect," *Mind*, vol. 70 (1961), p. 50.

* In his now overlooked *Logic*, W. E. Johnson points out the importance of seeing that apprehension is the conscious grasp of meaning.³¹ The "sensational processes are partly determined by past and possibly future experiences, and therefore cannot be wholly accounted for by the present sensations and contemporaneous neural processes" (p. 107). *Experiences*, not just neurology, play a causal role in present awareness. The intermingling of neural and mental processes is complex and not one-sided. Intention, for example, initiates new neural processes, although the person intending does not know what sort of neural processes are initiated, and although the processes are not what was intended (p. 108). To realize my intention, e.g., to pay back a debt, an effort has to be made to realize that intention. The consequences of this intention and effort are new neural processes, new physical effects of arms going out, money being exchanged, and an action: a debt repaid. The physical effects are inner and unknown (neural) and outer and foreknown (pp. 108-109). "As physical occurrences, the inner are causally determinative of the outer; but, in our analysis of the mental processes involved, we have to maintain, what may appear paradoxical, that it is the *foreknowledge* of the outer which causally determines the *occurrence* of the inner" (p. 109). Foreknowledge has "real causal efficiency." If mental processes, including foreknowledge, "could be reduced to merely physical or physiological terms, we should have to regard mental causality as an illusion" (p. 110). Johnson's grasp of the difference between mental and neural processes is sound. Among mental processes, he cites sensory, deliberative, and volitional processes. Within any one such process there may be causal relations, one desire may give rise to another, one thought leads to another. There are also causal interconnections across these various mental processes, as desire may influence my deliberations (p. 103). The flow of images is usually governed by association, but sometimes my interest modifies the course of association "and determines the flow of images or of ideas" (p. 103). Another instance of mental causation is the way attention can increase the determinateness of cognition. Such increased determinateness is the effort of attention.

Johnson does not overlook the role of the self in mental causation. When we attribute the activity of attention to the subject, we attribute "the

cause of the process to the agency of the subject" (p. 111). Earlier he had defined the active self as "a determinate phase of experience, which stands from time to time in definite and alterable relations to the processes that may be said to be actively controlled" (p. 104). The subject activities fall into two sorts, *motor* and *attentive*. The motor action relates the subject to the physical world, producing or preventing physical movement. For Johnson an important motor operation, one influenced by volitional or conative processes, is the effort at inhibiting a sneeze (p. 118). The attentive activities of the self relate to cognition alone. In both cases, Johnson says there is involved a more or less intense effort, motor effort and attentive effort. Since it seems doubtful whether any mental operations can occur without neural events, a question arises about the neural correlates of attentive effort, of cognition in general. Johnson suggests that the "activity of inner attention entails an operation upon the neural processes underlying *imagery*" (p. 112). The effort experienced would then be "due to the occurrence of imagery entailed in operating upon the neural processes" (p. 113). Further than this, Johnson does not wish to go in finding neural correlates for thinking. In fact, he asserts that there are no neural correlates to match the various sorts of cognitive states and processes. Besides the spatial and temporal order of neuronal impulses accompanying the perception of objects, what neural process goes along with the cognitive act of defining time? What neural processes accompany cognitive acts of comparison and relation? In general, "the reason why physiologists and psychologists never properly face the problem of the neural correlate of cognition, is because they virtually identify ideas with images" (p. 116). A striking example of the inadequacy of neurophysiology is given in our understanding of language. A sentence as understood "is not merely a whole for the thinker, but a *significant* whole." As Johnson remarks: "Mere association might give an adequate account of a combination of words which was mere nonsense; it cannot account for the added psychical fact that the sentence is understood as having meaning" (p. 117). What is the physiological process "correlated with the act of understanding" the significance of the sentence?

In cognition and understanding, the physical correlates of mental events disappear or fade into

³¹ Cambridge University Press, 1924. See especially his important chap. viii of Part III on "Application of Causal Notions to Mind."

the background conditions. "Thing seen" is not the physical stimuli involved in seeing, since seeing as apprehending lacks any separate and distinctive neurophysiological process. The second type of mental causation is sharply differentiated from the first. One is tempted to heighten this difference by saying, as Findlay does, that the causal factors, the inciting agents in such cases of mental causation, are the objects which "inexist" in thought.³² Such an intentional analysis of thought does not deny the physical conditions surrounding thought, it only stresses the mental features of mental causation. Findlay and other intentionalists become involved in a metaphysic of thought-objects, claiming that, since the objects which *inexist* in thought do not *exist*, "a distinguishing mark of the mental" is that "it alone can be made to act by what does not exist anywhere."³³

A metaphysic of such intentional objects is necessary, I think, for a comprehensive philosophy of mind; but even if we do not wish to go that far, Findlay makes several important observations about the causal relations in mental causation. Because a recognition of mental causes in this second sense is a recognition of non-physical causal factors, we are led into saying that future objects are causes of present thought and action. The Humean notion of cause works against such a recognition, but Findlay seems correct in saying "the *thought of* (or the *desire for*) the thing or consideration in question" is the cause of such mental activity.³⁴ The thought of some reason or motive can cause further *thought* or some *physical* action. The connection between the object thought of and the further thought or action is not contingent: "there is an essential affinity between the thought of an *A* and an *A* itself or *vice versa*, and there is more than merely 'verbal magic' in seeing the one fitly *continued* in the other."³⁵ Findlay is not very explicit about such an "essential affinity," though I think he does not have in mind the notion of formal objects of thought Kenny invokes.

Besides the notion of a cause being an intentional object inexistent or even future, another important difference between mental causation and Hume's notion of cause is the link between mental causation and knowledge without observation. Applying

this criterion of cause to mental activity has led some recent writers to look for special feelings always accompanying willing, intending, etc. Findlay points out the lack of any sharp contrast "between the *presence* of" such mental causation and "our *knowledge* of its presence."³⁶ We can "say how our experienced world will develop itself in and through our bodily movement, without *basing* such knowledge on our own previous thought and behaviour."³⁷ Just as I know without observation what I am doing, so I know without observation the causes of my doings and the intended results of my doings.

Knowing without observation the causes involved in mental causation is, of course, a function of my being the agent of my actions. It is the agency of such causation found especially, if not uniquely, in the third type of mental cause distinguished at the beginning of this section which requires that in analyzing causality we should return to force and influence. I think that Michotte is correct in showing that the ordinary sense of physical cause takes force, influence, and determination as central factors. Even if we bypass Hume and reintroduce such concepts into the notion of physical cause, the sort of force involved will not be the same as that of agent causation. A sense common to physical and mental cause may be that of something "growing out of" something else, of one thing "determining" another. The concept of cause as *initiator* of consequences may be basic.

V. KANT'S METAPHYSIC OF AGENCY

Kant has become famous for the way in which he adapted Hume's analysis of cause to fit a more general account of knowledge and the world. It has also been generally recognized that Kant saw the need for agent causality in his analysis of moral actions. But when he talks of cause in this latter context, he seems to violate his own injunction not to apply the categories beyond appearances. Kant found the need for a notion of cause other than the one he defined in the category sense. In dealing with the *ideas of reason*, Kant talks of applying the categories (at least, some of them)

³² J. N. Findlay, *Values and Intentions* (London, Allen and Unwin, 1961).

³³ *Ibid.*, p. 141.

³⁴ *Ibid.*, p. 142.

³⁵ *Ibid.*

³⁶ *Ibid.*

³⁷ *Ibid.*

not to the realm of freedom but to the totality of appearances. He speaks of the *conditioned* and the *unconditioned*. The concept of cause is associated with the idea of *origination* of the appearances.³⁸ When we think of the world as a unity in the existence of appearances, the condition of that which happens is called "cause." The "unconditioned causality in the field of appearance is called *freedom*, and its conditioned causality is called *natural cause* . . ." (p. 392). The unconditioned causality of appearances is freedom. It is important, too, to notice that "the world" here includes free causality. When we attempt to view the totality of appearance and to think of its beginnings, we find the very notion of "cause" which we need in thinking of free causality in action, cause as originator, and as initiator. The Third Antinomy underlines this important point.

If all the appearances occur in accordance with laws of nature, an infinite series of causes will occur without a beginning. The series of appearances will not be completed. The self-contradiction which the thesis of this Antinomy charges is due to this failure of completion of the series on the assumption that all events occur in accordance with laws of nature. The causality which we must assume to complete the series is described as being *absolute spontaneity* or transcendental freedom (p. 411). The antithesis shifts from "appearance" to "the world"; everything in the world takes place in accordance with laws of nature. Transcendental freedom is referred to as "a power of absolutely beginning a state" (p. 409). In denying the law of causality in this one instance, the antithesis "renders all unity of experience impossible" (p. 410). The point is also made that such freedom would be the absence of *guidance by rules* (p. 411). To think of freedom as determined in accordance with laws is to think of it as non-freedom. Since experience cannot be thought coherently without thinking it under rules, the way is blocked toward understanding freedom.

In his observations on this Antinomy, Kant says that in the cases of both kinds of causality we are unable to comprehend how one existence can be determined by another (p. 413). His example of an action stemming from free causality is that of his arising from his chair. The way he describes this action is important. A *new series* of events is initiated at that point, a series which has consequences even *in infinitum*. But Kant insists that

this new series is not a break with old events; is not a *beginning in time*, but only a *beginning in causality*. He significantly rephrases this distinction between causal beginning and temporal beginning as that between the *dynamically* first and the *mathematically* first, the latter term reinforcing the sense of "cause" as "ordered series." Both the resolution to arise and my act are a break causally with the natural events around this new series. The action *follows upon* the preceding events but does not *arise out of* them (p. 414). All the events in the world make up one time series, but within this time series there are a number of different causal series, some of which differ in their causality. More precisely, there is one causal series for the world, the time series, in which every part of the world has a place, but there are a number of initiating causes of different segments of this single time series.

While Kant maintains that we cannot understand how it is that one existence can be determined by another, he supplies us with some concise explanations of determination. Determination in the time series is a function of the position which any action or event has in that series. To see just how close to Leibniz Kant was on this matter, we need to add one other ingredient in our account of Kant's analysis of cause. This ingredient is his talk of two worlds—the sensible and the intelligible. Such talk was important in the *Grundlegung* and the practical *Critique*; it was clearly established also in the first *Critique*. Kant there speaks of some aspect of the sensible which is not sensible, of a faculty in the sensible appearance which is not itself the aspect of a sensible intuition. This faculty is the cause of appearances. For an understanding of this faculty we must form both a sensible and an intelligible concept of causality. The faculty refers, however, in both senses of causality to "one and the same effect" (p. 467). The causality of this faculty is intelligible (that is, non-sensible) while its effects are sensible. Corresponding to the dual notion of causality, there are two notions of the law of causality (Kant speaks of a "character" of causality). (1) There is the *empirical character* which refers to the connection between appearances, the uniform nature of appearances. Kant also speaks of the appearances (he uses the word "action" here) as being derived from other appearances. This interconnection of appearances constitutes "a single series in the

³⁸ References to Kant are to N. Kemp-Smith's translation. The discussion which follows should be read in conjunction with Section II of the paper referred to in n. 9 above.

order of nature" (p. 468). (2) There is the *intelligible character* which refers to something which is not appearance. This character is the cause of the same actions to which the empirical character refers. This cause of actions in the realm of appearance is the acting subject. In speaking of this acting subject it is inappropriate to say that action occurs, or to speak of time. Time is inapplicable to the subject as agent. There are no phenomena to be summarized in empirical laws. Since the acting subject as agent is not a phenomenon, it belongs to no empirical series. In short, "No action begins in this active being itself; but we may yet quite correctly say that the active being of itself begins its effects in the sensible world" (p. 469).

Two senses of "cause" are presented in the *Critique*. The one sense is that of lawful uniformity of appearances, the other sense is that of the initiation of events. The whole of Part III of the *Critique* is concerned with how reason can think the series of events which make up "nature" as having started, as having a beginning (cf., p. 475). It is the self as initiating cause which accounts for the beginning of many of the empirical series. Whether Kant also thought of extending this notion of initiating cause to the thing-in-itself is a question we cannot examine here. It is the self as agent, together with God, which in the final portions of the book provide the beginnings for the various series in nature or for the entire series. The "determination" of one event by another event, its predecessor, in the order of nature is not a causal determination in the ordinary sense. It is rather a logical, temporal, or serial determination. No member of the time series "could begin a series absolutely and of itself." Every event in that series is a link in the chain of nature (pp. 470-474). Even the actions which I, as agent, initiate are links in the chain of nature. The time-series notion of cause is clearly bound up with Leibnitz' notion of an ordered series, all positions of which are filled by prearranged plans. Kant even speaks of being able to predict with certainty all human actions (as appearances) from an exhaustive investigation of "all the appearances of men's wills" (p. 474). Reason "is the abiding condition of all those actions of the will under the guise of which man appears. Before even they have

happened, they are one and all predetermined in the empirical character" (p. 474). Reason as initiator gives rise to an entire series which makes up my experience, my life, a series which could only be changed if a different intelligible cause had initiated another series (p. 478).

VI. CONCLUSION

Kant's metaphysic of agency is an attempt to place human action in the world: to locate my world in the wider context of nature. I think it is the philosopher's task always to carry his analyses into ontology, whether he is engaged in the philosophy of action and of the person, or in epistemology. The ontologically faint of heart need not go that far. They must, however, find an understanding of causation sufficient for thought and action. For such an understanding we need to recognize that the relation between intentions, volitions, and emotions on the one hand, and action on the other, is causal, not logical; that this causal relation is inadequately analyzed in terms of correlation or expectation; and that the relations between the actions I do and myself is an especially intimate one best analyzed in terms like those of Merleau-Ponty, Vesey, and O'Shaughnessy.

The phenomenologists have explicated the meaning of "world," "experience," and "self," showing how these terms acquire a meaning for me in the course of my development. *The world* which I understand is *my world*, formed and developed by my action and my thought. The person becomes the center of a world elaborated in his thought, enacted through his actions. What is enacted is also known, though not through observation. Action has the modes of intending, willing, and deciding. Knowing has its modes of thinking, perceiving, and believing. The modes of action are also modes of knowing, and both are features of a self. The analyses of practical knowing and of doing made by phenomenologists and by analytical philosophers have revealed something of the nature of the person. A philosophy of the person is required to augment our understanding of action; that philosophy will have to say *what* a person is. A firmer grasp of the ontology of the agent will reinforce the analysis of agency which I have suggested.

III. NATURAL DEDUCTION RULES FOR OBLIGATION

FREDERIC B. FITCH*

I. INTRODUCTION

SOME systems of deontic logic have been proposed which take the deontic operators 'O' (obligation) and 'P' (permission) as definable in terms of some kind of implication together with a concept of the *bad situation* or the *violation*, V , so that 'Op' can be taken to mean that the denial of p implies V .¹ Other systems² take one or both of the operators 'O' and 'P' as undefined. On either approach it is natural to regard 'Op' as logically equivalent to ' $\sim P \sim p$ ' (where ' \sim ' denotes negation) and to regard 'Pp' as logically equivalent to ' $\sim O \sim p$ '.

Because of the close analogy between the logical properties of the obligation operation 'O' and those of the necessity operator ' \Box ', it seems worthwhile to formulate natural deduction rules for 'O' that are analogous to the natural deduction rules for ' \Box ' given in my book, *Symbolic Logic*. In many respects it is easier to work with such natural deduction rules than to employ a large set of axiom schemata and a single rule of procedure such as *modus ponens*.³ Also, the essential logical character of obligation (and permission) can be exhibited more clearly, or at least from another perspective, by constructing natural deduction rules specifically for the concept of obligation.

In my book I refer to the special kind of natural deduction procedures used there as constituting the *method of subordinate proofs*, since proofs appear within proofs in nested form.⁴ In the present paper this method of subordinate proofs will first be outlined before being applied to alethic modal logic or deontic modal logic. Then it will be applied to alethic modal logic, and finally to deontic modal logic. The method of subordinate proofs will be presented in somewhat more general form than

in my book, and the concept of *column* will be developed in such a way that a *proof* will be a special kind of column, and columns will be allowed to serve as hypotheses of other columns. This extension seems to add to the flexibility and usefulness of the method of subordinate proofs.

II. THE METHOD OF SUBORDINATE PROOFS

We begin by supposing that we have defined a class S of (*formal*) sentences. Of course S will be assumed to be such that the negation of a member of S is a member of S , and that the conjunction, disjunction, or implication of a member of S with a member of S is a member of S .

For convenience we will assume that the lower-case italic letters ' p ', ' q ', ' r ', and ' s ' are themselves sentences. In particular, ' $\sim p$ ' is the negation of ' p ', while ' $(p \& q)$ ', ' $(p \vee q)$ ', and ' $(p \supset q)$ ' are respectively the conjunction of ' p ' with ' q ', the disjunction of ' p ' with ' q ', and the implication (or conditional) of ' p ' with ' q '.

By a *sentential column* we mean a vertical line together with a vertically written sequence of sentences to the right of it. For example, the following expression is a sentential column:

	p
	q
	$r \& s$
	$\sim q$
	q
	$\sim(s \supset p)$

* This paper was written in early 1964 and was submitted for publication in July of that year. (See also the author's abstract in the *Journal of Symbolic Logic*, vol. 29 [1964], pp. 150-151, entitled "Natural Deduction Rules for Obligation.")

¹ A. R. Anderson, "A Reduction of Deontic Logic to Alethic Modal Logic," *Mind*, vol. 67 (1958), pp. 100-103.

² For example, G. H. von Wright, *An Essay in Modal Logic* (Amsterdam, 1951).

³ Nevertheless the use of natural deduction procedures is often equivalent, in an important sense, to use of axiom schemata and *modus ponens*. This is true in the case of the natural deduction procedures of the present paper.

⁴ This nested arrangement of proofs seems to be due originally to S. Jaśkowski, "On the Rules of Suppositions in Formal Logic," *Studia Logica*, no. 1 (Warsaw, 1934).

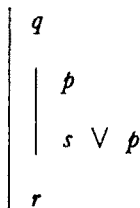
Notice that the outer parentheses of ' $(r \ \& \ s)$ ' were omitted. It will be customary to omit outer parentheses when ambiguity does not result from doing so.

Just as we may distinguish between a sentence and an occurrence of a sentence, so also, of course, we may distinguish between a sentential column and an occurrence of a sentential column. This sort of distinction, even when not overtly made, is usually clear enough from the context when required, so we shall not bother to emphasize it. A similar distinction is also to be understood in the case of the concepts of *item* and *column*, next to be defined; that is, there is to be understood to be a distinction between an item and an occurrence of an item, and a distinction between a column and an occurrence of a column.

The concepts of *item* and *column* are defined by a simultaneous induction as follows:

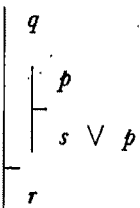
- (1) Every sentence (member of S) is an *item*.
- (2) A vertical line together with a vertically written sequence of *items* to the right of it is a *column* (and if the items are all of them sentences, then the column is a sentential column).
- (3) Every *column* is an *item*.
- (4) The only *items* and *columns* are those that are such in virtue of (1)–(3).

For example,



is a column having three items, the first of which is the sentence ' q ', the second of which is the sentential column that has ' p ' and ' $s \vee p$ ' as its own items, and the third of which is the sentence ' r '.

Some items of a column may be designated as being *hypotheses* of that column. The items thus designated, if there are any, must always be grouped together at the top of the column and separated from the other items of the column by a short horizontal line extending out to the right of the vertical line. Thus,



is a column just like the previous example except that the first two items (the sentence ' q ' and the two-item sentential column) have been designated as hypotheses of the main column, while the first item of the two-item sentential column has been designated as a hypotheses of that column.

Every item of a column will be said to be *directly subordinate* to the column. Thus, in the above example, the sentence ' q ', the two-item sentential column, and the sentence ' r ' are directly subordinate to the main column. On the other hand, the sentences ' p ' and ' $s \vee p$ ' are *not* directly subordinate to the main column, though they are of course directly subordinate to the two-item sentential column, being the two items of it.

The concept of *subordination* may now be defined inductively in terms of the concept of direct subordination, as follows:

- (1) If an item is directly subordinate to another item, then it is *subordinate* to that other item.
- (2) If an item is *subordinate* to another item, and if that other item is *subordinate* to a third item, then the first item is *subordinate* to the third item.
- (3) An item is subordinate to another item only if its being so follows from (1) and (2).

Clearly subordination is transitive but not reflexive or symmetrical. In the example given above, the sentences ' q ', ' p ', ' $s \vee p$ ', and ' r ', and the two-item sentential column are all subordinate to the main column, though, as already noted, ' p ' and ' $s \vee p$ ' are not directly subordinate to it. Thus we will say that ' p ' and ' $s \vee p$ ' are *indirectly subordinate* to the main column. The sentence ' s ', moreover, is not subordinate to either the main column or the shorter two-item sentential column, but it is merely part of the sentence ' $s \vee p$ ' which is subordinate to both these columns.

We assume, next, that a concept of *direct consequence* has been defined, according to which a sentence may be said to be a direct consequence of one or more preceding items of the same column. For example, it would be reasonable to assume that ' $p \ \& \ q$ ' is a direct consequence of ' p ' and ' q ', and that ' $p \vee q$ ' is a direct consequence of ' p '. We may also assume, if we wish, that some sentences have been specified as being *axioms*. For example, ' $p \vee \sim p$ ' might be so specified. Finally, we suppose that a relation of *reiteration* has been defined, according to which various sentences may be said to be *reiterates* of other sentences or of themselves. Usually each sentence is treated as the only reiterate of itself, but in modal and deontic logic there are exceptions to this, as we shall see

later. For the present we shall assume that each sentence is the only reiterate of itself. The use of the concept of reiteration will be explained in connection with the definition of the concept of *proof*, to which we now turn.

By a *proof* we mean a column C such that each item i of C (that is, each item i directly subordinate to C) satisfies at least one of the following conditions:

- (1) i is a hypothesis of the column C .
- (2) i is an axiom.
- (3) i is a direct consequence of preceding items of the column C (that is, of items which have occurrences as directly subordinate to C and which, in those occurrences, precede i).
- (4) i is itself a proof.
- (5) i is a reiterate of an item j of a column D to which column C is directly subordinate, and j (in its occurrence as a member of the column D) precedes the column C . (Thus j and C both occur as directly subordinate to D , and these occurrences are such that the occurrence of j precedes that of C , and indeed precedes all such occurrences of C , if there is more than one. Furthermore, i and j will be the same sentence, since, at the outset, we are assuming that the only reiterate of a sentence is that sentence itself.)

For example, if ' $p \& q$ ' is assumed to be a direct consequence of ' p ' and ' q ', if ' $s \vee (p \& q)$ ' is assumed to be a direct consequence of ' $p \& q$ ', and if ' $q \supset (s \vee (p \& q))$ ' is treated as being a direct consequence of the column,

	q
	p
	$p \& q$
	$s \vee (p \& q)$

then the following is a proof:

1	p	hypothesis of column 1-6,
2	q	hypothesis of column 2-5,
3	p	reiterate of 1,
4	$p \& q$	direct consequence of 2 and 3,
5	$s \vee (p \& q)$	direct consequence of 4,
6	$q \supset (s \vee (p \& q))$	direct consequence of column 2-5.

* W. V. Quine, *Mathematical Logic* (New York, 1940).

The numbers on the left have been added to facilitate reference to various parts of the proof, and the remarks on the right show that column 1-6 does satisfy the requirements for being a proof. Notice that the subordinate column 2-5 is also a proof but would not be so if removed from the context contributed by the larger proof 1-6, since the second item of 2-5 would not then satisfy any of the conditions (1)-(5) listed above.

We now introduce the following rules which serve to define the relation of *direct consequence*. For convenience we use square corners and Greek letters as Quine does, except that we place square corners around isolated Greek letters, whereas Quine does not.⁵ We let "d.c." serve as an abbreviation for the phrase, "direct consequence."

Rules of direct consequence:

' $\phi \& \psi$ ' is a d.c. of the pair of sentences ' ϕ ' and ' ψ ' by *conjunction introduction*.

' ϕ ' and ' ψ ' are d.c.s of ' $\phi \& \psi$ ' by *conjunction elimination*.

' $\phi \vee \psi$ ' is a d.c. of ' ϕ ' by *disjunction introduction*, and it is also a d.c. of ' ψ ' by *disjunction introduction*.

' θ ' is a d.c., by *disjunction elimination*, of ' $\phi \vee \psi$ ' and the columns

ϕ	ψ
.	.
.	.
.	.
θ	θ

(The triple dots indicate the possible presence of other items in the columns.)

' ϕ ' is a d.c. of ' $\sim\sim\phi$ ' by *double negation elimination*.

' $\sim\phi$ ' is a d.c., by *negation introduction*, of any column which has ' ϕ ' as its only hypothesis and which contains a pair of contradictory sentences ' ψ ' and ' $\sim\psi$ ' among its items.

' ϕ ' is a d.c. of any pair of contradictory sentences ' ψ ' and ' $\sim\psi$ ' by *negation elimination*.

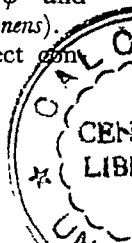
' $\phi \supset \psi$ ' is a d.c. of the column

ϕ
.
.
.
ψ

by *implication introduction*.

' ψ ' is a d.c. of the pair of sentences ' ϕ ' and ' $\phi \supset \psi$ ' by *implication elimination (modus ponens)*.

In addition to the above rules of direct consequence



sequence we also assume that every sentence can serve as a reiterate of itself.

The above rules provide a formalization of the two-valued propositional calculus. For example, a proof of ' $p \vee \sim p$ ' can be constructed as follows:

1	$\sim(p \vee \sim p)$	hypothesis of column 1-6,
2	p	hypothesis of column 2-4,
3	$p \vee \sim p$	d.c. of 2 by disjunction introduction,
4	$\sim(p \vee \sim p)$	reiterate of 1,
5	$\sim p$	d.c. of column 2-4 by negation introduction,
6	$p \vee \sim p$	d.c. of 5 by disjunction introduction,
7	$\sim\sim(p \vee \sim p)$	d.c. of column 1-6 by negation introduction,
8	$p \vee \sim p$	d.c. of 7 by double negation elimination.

The above proof is said to be a *categorical proof* because it has no hypotheses among its items (that is, there is no item which occurs as a hypothesis directly subordinate to it), though there are some hypotheses which are indirectly subordinate to it. Here is another example of a proof:

1	$p \supset r$	first hypothesis of column 1-11,
2	$q \supset r$	second hypothesis of column 1-11,
3	$p \vee q$	hypothesis of column 3-10,
4	p	hypothesis of column 4-6,
5	$p \supset r$	reiterate of 1,
6	r	d.c. of 4 and 5 by implication elimination,
7	q	hypothesis of column 7-9,
8	$q \supset r$	reiterate of 2,
9	r	d.c. of 7 and 8 by implication elimination,
10	r	d.c. of 3, 4-6, and 7-9 by disjunction elimination,
11	$(p \vee q) \supset r$	d.c. of column 3-10 by implication introduction.

The above proof has hypotheses, and so it is said to be a *hypothetical proof* rather than a categorical proof.

It is perhaps worth remarking that the three

rules, double negation elimination, negation elimination, and negation introduction, can be replaced by a single rule of *indirect proof*, as follows:

' ϕ ' is a d.c., by *indirect proof*, of any column which has ' $\sim\phi$ ' as its only hypothesis and which contains a pair of contradictory sentences ' ψ ' and ' $\sim\psi$ ' among its items.

This latter rule could have been used in the proof of ' $p \vee \sim p$ ' given previously, since, in that proof, it is seen that ' $p \vee \sim p$ ' is a direct consequence of column 1-6 by indirect proof. In fact, it is possible to rewrite that whole proof so that the only rules used are disjunction introduction and indirect proof (but also allowing reiteration, of course).

If the rule of double negation elimination is omitted, and also the rule of indirect proof (since the latter gives the effect of the rule of double negation), then the remaining rules so far described give rise to the Heyting system of intuitionistic logic⁶ formulated in subordinate-proof style.

Other examples of proofs can be found in my book, *Symbolic Logic*. But the method of subordinate proofs, as described in my book, is somewhat less general than the method described above. The method described above (though not the method of my book) allows us to add two further rather interesting and useful rules of direct consequence for columns themselves, as follows:

If one column has exactly the same hypotheses as another column, but lacks one or more other items that the other column has, then it is a d.c. of that other column by *column introduction*.

' ϕ ' is a d.c., by *column elimination*, of ' ψ_1 ', ' ψ_2 ', ..., ' ψ_n ' and a column having ' ψ_1 ', ' ψ_2 ', ..., ' ψ_n ' as its only hypotheses and having ' ϕ ' as one of its items.

We will assume in the above two rules and hereafter that Greek letters may refer to columns as well as to sentences when the context allows such reference as meaningful. In particular, there is nothing that precludes the various items referred to in the above two rules from being columns rather than sentences. Furthermore, we will assume that every column, as well as every sentence is a reiterate of itself. Thus it will be permissible to reiterate columns into other columns in the same way that sentences can be reiterated into columns. Also, columns may themselves be hypotheses of other columns, so that the hypotheses referred to in the above two rules might themselves in some

⁶ See A. Heyting, *Intuitionism* (Amsterdam, 1956). For an intuitionistic system of modal logic, see F. B. Fitch, "Intuitionistic Modal Logic with Quantifiers," *Portugaliae Mathematica*, vol. 7 (1948), pp. 113-118.

cases be columns. Thus not all columns that appear in proofs are themselves proofs. Some non-proof columns may be present in the role of hypotheses for other columns. Other non-proof columns may be present as direct consequences of preceding columns. This situation contrasts with the situation in my book, where all columns appearing anywhere in a proof are themselves proofs. We now consider some examples of proofs in which the rules of column introduction and column elimination are used.

1	p	first hypothesis of column 1-7,
2	p	} second hypothesis of column 1-7,
3	q	
4	r	
5	q	d.c. of 1 and column 2-4 by column elimination,
6	r	d.c. of 1 and column 2-4 by column elimination,
7	$q \& r$	d.c. of 5 and 6 by conjunction introduction.

1	p	} hypothesis of column 1-5,
2	q	
3	r	
4	p	} d.c. of column 1-3 by column introduction.
5	r	

1	p	} hypothesis of column 1-10,
2	q	
3	p	first hypothesis of column 3-7,
4	r	second hypothesis of column 3-7,
5	p	} reiterate of column 1-2,
6	q	
7	q	d.c. of 3 and column 5-6 by column elimination,
8	p	} d.c. of column 3-7 by column introduction.
9	r	
10	q	

The following example is somewhat like the above example, but it uses the case of column elimination where $n = 0$. We assume that this case of the rule is available.

1	q	} column 1-1, hypothesis of column 1-6,
2	r	
3	q	} column 3-3, reiterate of column 1-1,
4	q	
5	r	} d.c. of column 2-4 by column introduction.
6	q	

It is easy to give a proof of the column

p
r

from the hypotheses

p	and	q
q		r

This is left

as an exercise for the reader. Many other interesting properties of columns can also be proved.

As a final example, we give a demonstration that ' $\sim p$ ' follows from a column that has ' p ' as its only hypothesis and has ' q ' and ' $\sim q$ ' as its only other items. Since the rule of negation introduction is not used in this proof, the proof amounts, in a sense, to a derivation of the rule of negation introduction from the rule of indirect proof and the column rules.

1	p	} hypothesis of column 1-13,
2	q	
3	$\sim q$	
4	$\sim \sim p$	hypothesis of column 4-12,
5	$\sim p$	hypothesis of column 5-6,
6	$\sim \sim p$	reiterate of 4,
7	p	d.c. of column 5-6 by indirect proof,
8	p	} reiterate of column 1-3,
9	q	
10	$\sim q$	
11	q	d.c. of 7 and column 8-10 by column elimination,
12	$\sim q$	d.c. of 7 and column 8-10 by column elimination,
13	$\sim p$	d.c. of column 4-12 by indirect proof.

An additional rule of direct consequence that is not really needed, but sometimes seems convenient,

is the *rule of repetition* that states that each sentence (and each column, too, if we wish) is a d.c. of itself.

It should also be remarked that the process of reiteration can be made slightly more flexible than described here. This can be done by providing for reiterations into columns that are indirectly subordinate to the column of the reiterated item.

III. SUBORDINATE PROOFS IN MODAL LOGIC

In modal logic it is usual to employ an operator ' \Box ' standing for *necessity*, so that ' $\Box p$ ' would be read as " p is necessary." We assume in modal logic that the result of operating on any member of S with the operator ' \Box ' is again a member of S , where S is the class of (formal) sentences.

When subordinate-proof method is being used in modal logic, it is convenient to assume not only that we have columns in the sense already described (and they may be called *regular columns*) but also columns of a special kind which may be called *strict columns*. A strict column is indicated by placing the symbol ' \Box ' to the immediate left of the upper part of the vertical line of the column. Here is an example of a strict column:

\Box	p
	q
	r

All the rules of direct consequence so far described can be used also in modal logic, but wherever these rules mention columns, we are to understand that the reference is to regular columns, not strict columns. On the other hand, the conditions (1)–(5), which determine when a column C is a proof, are equally applicable to the case where C is a strict column as they are to the case where C is a regular column, except that condition (5) for a *strict* column C requires a concept of *strict reiteration* and reads just the same except that the concept of strict reiteration replaces the concept of reiteration.

For the subordinate-proof versions of the various modal systems M, B, S₄, and S₅,⁷ the corresponding concepts of strict reiteration are defined as

follows: For all the systems M, B, S₄, and S₅, ' $\Box\phi$ ' is a strict reiterate of ' $\Box\phi$ '. For the systems S₄ and S₅, ' $\Box\phi$ ' is a strict reiterate of itself. For the systems B and S₅, ' $\Box\phi$ ' is the strict reiterate of ' $\sim\phi$ '. (An alternative definition for strict reiteration in the system S₅ would assert that the only strict reiterates in the system S₅ are as follows: ' ϕ ' is the strict reiterate of ' $\Box\phi$ ', while ' $\sim\Box\phi$ ' is the strict reiterate of itself. This different way of defining strict reiteration for S₅ would change the form of proofs in S₅, but the same theorems would be provable.)

All four of these modal systems can be regarded as employing the following two rules of direct consequence for necessity:

' $\Box\phi$ ' is a d.c., by *necessity introduction*, of any strict column that has no hypotheses and has ' ϕ ' as an item.

' ϕ ' is a d.c. of ' $\Box\phi$ ' by *necessity elimination*.

Here is an example of a column which is a proof in all four modal systems M, B, S₄, and S₅:

1	$\Box p$	first hypothesis of column 1–6,
2	$\Box q$	second hypothesis of column 1–6,
3	$\Box p$	strict (M-, B-, S ₄ -, S ₅ -)reiterate of 1,
4	q	strict (M-, B-, S ₄ -, S ₅ -)reiterate of 2.
5	$p \& q$	d.c. of 3 and 4 by conjunction introduction,
6	$\Box(p \& q)$	d.c. of strict column 3–5 by necessity introduction.

Here are three other columns which also are proofs in the four systems M, B, S₄, and S₅:

1	$\Box(p \& q)$	hypothesis of column 1–7,
2	$\Box p \& q$	strict (M-, B-, S ₄ -, S ₅ -) reiterate of 1,
3	p	d.c. of 2 by conjunction elimination,
4	q	d.c. of 2 by conjunction elimination,
5	$\Box p$	d.c. of strict column 2–4 by necessity introduction,
6	$\Box q$	d.c. of strict column 2–4 by necessity introduction,
7	$\Box p \& \Box q$	d.c. of 5 and 6 by conjunction introduction.

⁷ For the system M, see von Wright, *op. cit.* For the systems S₄ and S₅, see C. I. Lewis and C. H. Langford, *Symbolic Logic* (New York, 1932), Appendix II. See also, A. N. Prior, *Formal Logic* (London, 1955). The system we here call B is usually referred to as the *Brouwersche* system. This system and the systems M, S₄, and S₅ are discussed by S. A. Kripke, "Semantical Considerations in Modal Logic," *Acta Philosophica Fennica*, vol. 16 (1963), pp. 83–94. Kripke has inspired some of the ideas of the present paper.

1		$\Box\Box p$	hypothesis of column 1-2,
2		$\Box p$	d.c. of 1 by necessity elimination.
1		$\Box(p \supset q)$	hypothesis of column 1-8,
2		$\Box p$	hypothesis of column 2-7,
3		$\Box(p \supset q)$	reiterate of 1,
4		$\Box p \supset q$	strict (M-, B-, S4-, S5-)reiterates of 3,
5		p	strict (M-, B-, S4-, S5-)reiterate of 2,
6		q	d.c. of 4 and 5 by implication elimination,
7		$\Box q$	d.c. of strict column 4-6 by necessity introduction,
8		$\Box p \supset \Box q$	d.c. of column 2-7 by implication introduction.

Notice that at 3 in the above proof we have an ordinary reiterate, and recall that every sentence is an ordinary reiterate of itself. If a column, such as 2-7 above, is an ordinary column, then the reiterates that appear as items of it should be ordinary reiterates. On the other hand, column 4-6 above is a strict column, so the two reiterates that appear as items of it must be strict reiterates. Here is an example of a column that is a proof in the systems S4 and S5, but not in the system B or M:

1		$\Box p$	hypothesis of column 1-3,
2		$\Box \Box p$	strict (S4-, S5-)reiterate of 1, and the only item of 2-2,
3		$\Box\Box p$	d.c. of strict column 2-2 by necessity introduction.

The above column would not be a proof in M or B because ' $\Box p$ ' is not a strict reiterate of itself in either of those systems. Here is an example of a column that is a proof in the systems B and S5, but not in the systems M or S4 owing to the fact that ' $\sim\Box p$ ' is not a strict reiterate of ' $\sim p$ ' in M or S4.

1		$\sim p$	hypothesis of column 1-3,
2		$\Box \sim\Box p$	strict (B-, S5-)reiterate of 1, and the only item of 2-2,
3		$\Box\sim\Box p$	d.c. of strict column 2-2 by necessity introduction.

The main difference of the systems S4 and S5 from the systems M and B is that ' $\Box\Box\phi$ ' is derivable from ' $\Box\phi$ ' in the former two systems but not in the latter two systems, while the main

difference of the systems B and S5 from M and S4 is that ' $\Box\sim\Box\phi$ ' is derivable from ' $\sim\phi$ ' in B and S5, but not in M or S4. Also, of these four systems, S5 is the only one in which ' $\Box\sim\Box\phi$ ' is derivable from ' $\sim\Box\phi$ '.

In modal logic the concept of possibility is often represented by the symbol ' \Diamond '. Thus ' $\Diamond p$ ' means that p is possible, just as ' $\Box p$ ' means that p is necessary. In systems of modal logic such as M, B, S4, and S5 which are extensions of two-valued propositional logic, we may define possibility in terms of necessity by treating ' $\Diamond p$ ' as an abbreviation for ' $\sim\Box\sim p$ ', and in general treating ' $\Diamond\phi$ ' as an abbreviation for ' $\sim\Box\sim\phi$ '. If this is done, then the following four rules become derivable (or "admissible") in the sense that adding them would not make it possible to prove any sentence not already provable without them:

' $\Diamond\phi$ ' is a d.c. of ' ϕ ' by *possibility introduction*.

' $\Diamond\psi$ ' is a d.c., by *possibility elimination*, of ' $\Diamond\phi$ ' and a strict column that has ' ϕ ' as its only hypothesis and that has ' ψ ' as one of its items.

' $\sim\Diamond\phi$ ' is a d.c. of ' $\Box\sim\phi$ ' by *negative possibility introduction*.

' $\Box\sim\phi$ ' is a d.c. of ' $\sim\Diamond\phi$ ' by *negative possibility elimination*.

The following proof, for example, shows that ' $\Diamond p$ ' is derivable from ' p ', so that the effect of the rule of possibility introduction can be obtained without assuming that rule itself:

1		p	hypothesis of column 1-5,
2		$\Box\sim p$	hypothesis of column 2-4,
3		$\sim p$	d.c. of 2 by necessity elimination,
4		p	reiterate of 1,
5		$\Diamond p$	d.c. of column 2-4 by negation introduction, and definition.

If the last item of the above column is written in its abbreviated form as ' $\sim\Box\sim p$ ', the fact that it is a direct consequence of the subordinate column 2-4 by negation introduction should be clear.

The introduction and elimination rules for columns can be included among the rules of the systems M, B, S4, and S5 without thereby adding to the class of provable sentences of any of these systems. If this is done, then it would also be natural to extend the rules of necessity introduction and necessity elimination to cases where ' ϕ ' is a column instead of a sentence. This can be done

by understanding ' $\Box\phi$ ' to be the strict column obtained by prefixing ' \Box ' to ' ϕ '. Furthermore, in all four modal systems, we could define the strict reiterate of a strict column ' $\Box\phi$ ' as being simply the regular column ' ϕ ' itself. The class of sentences provable in these systems would not be affected by these additional procedures. Here are two proofs that illustrate these procedures with columns:

1	$\Box p$	first hypothesis of column 1-8,
2	p	} second hypothesis of column 1-8,
3	q	
4	$\Box p$	strict (M-, B-, S ₄ -, S ₅ -)reiterate of 1,
5	p	} strict (M-, B-, S ₄ -, S ₅ -)reiterate of column 2-3,
6	q	
7	q	d.c. of 4 and column 5-6 by column elimination,
8	$\Box q$	d.c. of column 4-7 by necessity introduction.

1	$\Box p$	} hypothesis of column 1-10,
2	q	
3	r	
4	$\Box p$	} strict reiterate of column 1-3,
5	q	
6	r	
7	p	} d.c. of column 4-5 by column introduction.
8	r	
9	$\Box p$	} d.c. of column 4-8 by necessity introduction.
10	q	

As another example, it could be shown that ' $\Diamond q$ ' is provable from the hypotheses ' $\Diamond p$ ' and the strict column that has ' p ' as its only hypothesis and ' q ' as its only other item. Since this can be done without using the rule of possibility elimination, it amounts, in a sense, to a derivation of the rule of possibility elimination from the other available rules, including the rules for columns.

Without presenting the details here, it is worth noting that all the systems so far considered can easily be extended to deal with the universal and

existence quantifiers. This can be done by suitably redefining the class S of formal sentences and by employing introduction and elimination rules for the universal and existence quantifiers similar to the rules for them used in my book. In this way we obtain the quantified modal systems QM, QB, QS₄, and QS₅. We get slightly different forms of these quantified modal systems depending on whether or not we treat ' $\Box(a)\phi$ ' as being the strict reiterate of ' $\Box(a)\phi$ '. If we do so, then the "Barcan formula," ' $\Box(a)\phi \supset \Box(a)\phi$ ', is derivable.³

IV. SUBORDINATE PROOFS IN DEONTIC LOGIC

The systems of modal logic that we have so far been considering are said to be *alethic* systems (from the Greek word for "truth") because they are concerned with necessity in the sense of *necessary truth*. Thus in all these systems, if a proposition is necessary, it is true. There exist other modal systems, however, which contain a concept of necessity for which this is not the case, namely, *obligation*, that is, *ethical* or *legal necessity*. Since not all obligations are fulfilled, propositions can be necessary in this sense without being true. Systems of modal logic that deal with the concept of obligation are called *deontic* systems, from the Greek word for "ought."

The simplest way to convert the alethic modal systems M, B, S₄, and S₅ into corresponding deontic modal systems DM, DH, DS₄, and DS₅, is to replace the rule of necessity elimination by the following rule of *weak necessity elimination*:

' $\Diamond\phi$ ' is a d.c. of ' $\Box\phi$ ' by *weak necessity elimination*. We assume here that the possibility operator, ' \Diamond ', has been defined in the way previously indicated. Not only must the rule of necessity elimination be replaced by the rule of weak necessity elimination, but a change must also be made in the concept of strict reiteration in the case of the systems DB and DS₅. This change will be specified shortly. The rule of possibility introduction will no longer be derivable in the deontic systems, but the other possibility rules previously mentioned will remain derivable.

In the various deontic systems we will use the symbol 'O' in place of the symbol ' \Box ', and the symbol 'P' in place of the symbol ' \Diamond '. We can consider 'O' as referring to *obligation* and 'P' as referring to *permission*. Strict columns will be called *obligative columns*, and they will be indicated by

³ Ruth C. Barcan (Mrs. J. A. Marcus), "A Functional Calculus of First Order Based on Strict Implication," *Journal of Symbolic Logic*, vol. 29 (1964), pp. 1-16.

prefixing an 'O' instead of a ' \square '. The two rules for obligation are then as follows, being simply restatements of the rule of necessity introduction and the rule of weak necessity elimination:

' $O\phi$ ' is a d.c., by *obligation introduction*, of any obligative column that has no hypotheses and has ' ϕ ' as an item.

' $P\phi$ ' (that is, ' $\sim O\sim\phi$ ') is a d.c. of ' $O\phi$ ' by *obligation elimination*.

In the deontic systems we will speak of *obligative reiterates* instead of strict reiterates. In all the deontic systems ' ϕ ' will be an obligative reiterate of ' $O\phi$ '. In DS_4 and DS_5 , ' $O\phi$ ' will be an obligative reiterate of itself, while in DB and DS_5 , ' $\sim O\phi$ ' will be the obligative reiterate of itself. We retain an exact analogy with M and S_4 in the case of DM and DS_4 as far as obligative reiterates are concerned, but we do not retain such an exact analogy with B and S_5 in the case of DB and DS_5 , for if we did so we would have to treat ' $\sim O\phi$ ' as being an obligative DB- and DS_5 -reiterate of ' $\sim\phi$ ', and this would lead to the odd result, ' $\sim\phi \supset O\sim O\phi$ ', and in the case of DS_5 even to the result, ' $O\phi \supset \phi$ ', thereby reinstating the effect of necessity elimination and converting DS_5 into an alethic system equivalent to S_5 itself. Therefore in DB and in DS_5 we treat ' $\sim O\phi$ ' as the obligative reiterate of itself rather than the obligative reiterate of ' $\sim\phi$ '.

We previously gave a proof of ' $\square(p \& q)$ ' from the hypotheses ' $\square p$ ' and ' $\square q$ '. This was a correct proof in all the alethic systems M, B, S_4 , and S_5 . By merely trivial changes, this proof can be converted into a proof of ' $O(p \& q)$ ' from the hypotheses ' $O p$ ' and ' $O q$ ', as follows:

1	$O p$	first hypothesis of column 1-6,
2	$O q$	second hypothesis of column 1-6,
3	$O \mid p$	obligative (DM-, DB-, DS_4 -, DS_5 -)reiterate of 1,
4	q	obligative (DM-, DB-, DS_4 -, DS_5 -)reiterate of 2,
5	$p \& q$	d.c. of 3 and 4 by conjunction introduction,
6	$O(p \& q)$	d.c. of obligative column 3-5 by obligation introduction.

The above column is a proof in all the deontic systems DM, DB, DS_4 , and DS_5 . Similarly, just as previously we gave a proof of ' $\square p \& \square q$ ' from the hypothesis ' $\square(p \& q)$ ', so also we can give a proof of ' $O p \& O q$ ' from the hypothesis ' $O(p \& q)$ ', and it will be a correct proof in all the deontic

systems that we are considering. Also, in the same way, a proof of ' $O p \supset O q$ ' can be given from the hypothesis ' $O(p \supset q)$ ', as follows:

1	$O(p \supset q)$	hypothesis of column 1-8,
2	$O p$	hypothesis of column 2-7,
3	$O(p \supset q)$	reiterate of 1,
4	$O \mid p \supset q$	obligative (DM-, DB-, DS_4 -, DS_5 -)reiterate of 3,
5	p	obligative (DM-, DB-, DS_4 -, DS_5 -)reiterate of 2,
6	q	d.c. of 4 and 5 by implication elimination,
7	$O q$	d.c. of 4-6 by obligation introduction,
8	$O p \supset O q$	d.c. of 2-7 by implication introduction.

Similarly, a proof of ' $O O p \supset O p$ ' can be given from the hypothesis, ' $O(O p \supset p)$ '. Furthermore, as we shall see, the sentence ' $O(O p \supset p)$ ' is itself provable in the systems DB and DS_5 , so that the sentence ' $O O p \supset O p$ ' is also provable in both those systems, though neither ' $O(O p \supset p)$ ' nor ' $O O p \supset O p$ ' is provable in DM or in DS_4 . Here is a proof of ' $O(O p \supset p)$ ' in the systems DB and DS_5 :

1	$O p \vee \sim O p$	excluded middle (proved as in proof of ' $p \vee \sim p$ '),
2	$O p$	hypothesis of column 2-7,
3	$O \mid p$	obligative (DB-, DS_5 -)reiterate of 2,
4	$O p$	hypothesis of column 4-5,
5	p	reiterate of 3,
6	$O p \supset p$	d.c. of column 4-5 by implication introduction,
7	$O(O p \supset p)$	d.c. of column 3-5 by obligation introduction,
8	$\sim O p$	hypothesis of column 8-14,
9	$O \mid \sim O p$	obligative (DB-, DS_5 -)reiterate of 8,
10	$O p$	hypothesis of column 10-12,
11	$\sim O p$	reiterate of 9,
12	p	d.c. of 10 and 11 by negation elimination,
13	$O p \supset p$	d.c. of column 10-12 by implication introduction,
14	$O(O p \supset p)$	d.c. of column 9-13 by obligation introduction,
15	$O(O p \supset p)$	d.c. of 1, 2-7, 8-14 by disjunction elimination.

For the above column to be technically a proof, the proof of ' $Op \vee \sim Op$ ' would have to be inserted at the beginning. This would of course proceed in the same way as the proof of ' $p \vee \sim p$ ' given earlier. The resulting total proof would be fairly long, but a shorter total proof can be given as presented below (omitting explanatory comments). Although ' $OOp \supset Op$ ' can be proved using ' $O(Op \supset p)$ ', in the way previously indicated, a more direct proof of ' $OOp \supset Op$ ' is given below (again omitting explanatory comments). These are proofs in DB and DS₅, but not in DM or DS₄.

1	$\sim O(Op \supset p)$	1	OOp
2	$O \mid \sim O(Op \supset p)$	2	$\sim Op$
3	Op	3	OOp
4	$O \mid p$	4	$O \mid \sim Op$
5	Op	5	Op
6	p	6	p
7	$Op \supset p$	7	Op
8	$O(Op \supset p)$	8	Op
9	$\sim O(Op \supset p)$	9	$OOp \supset Op$
10	p		
11	$Op \supset p$		
12	$O(Op \supset p)$		
13	$O(Op \supset p)$		

Notice that in all the alethic modal systems ' $\Box p$ ' follows from ' $\Box \Box p$ ' merely by necessity elimination, so we easily get ' $\Box \Box p \supset \Box p$ ' by use of implication introduction. The corresponding deontic sentence, ' $OOp \supset Op$ ', is not provable in all the deontic systems, and where provable, the proof is more difficult. Similar remarks apply to ' $\Box(\Box p \supset p)$ ' and ' $O(Op \supset p)$ '.

Just as the introduction and elimination rules for columns could be included among the rules for the alethic modal systems, the same is true for the deontic modal systems. We can treat each column as the (ordinary) reiterate of itself and each regular column as the obligative reiterate of the obligative column obtained by prefixing an 'O' to it. We can also regard the rule of obligation introduction as applying to columns as well as to sentences in

the same way as the rule of necessity introduction can be applied to columns. On the other hand, the rule of obligation elimination, from its very nature, cannot be applied to columns in the way that the rule of necessity elimination can.

V. DEONTIC INCORRIGIBILITY

The various deontic systems DM, DB, DS₄, and DS₅ provide, as it were, various different theories of obligation, or theories of different kinds of obligation. These seem to be among the simplest and most interesting deontic systems, though others can of course easily be devised. For example, we could get a modified form of DM by adding to it as an axiom each sentence ' $O(Op \supset p)$ '. The system DS₄ could be modified in a similar way. The weakest of these various systems is DM. It has the advantage of asserting fewer properties of obligation than the others and so, perhaps, of being less likely to be in error. In particular it does not assert that if p is obligatory then p ought to be obligatory (i.e., if p is obligatory, then it is obligatory that p is obligatory), while DS₄ and DS₅ do assert this, as is seen as follows:

1	Op	hypothesis of column 1-3,
2	$O \mid Op$	obligative (DS ₄ , DS ₅)-reiterate of 1,
3	OOp	d.c. of obligative column 2-2 by obligation introduction.
4	$Op \supset OOp$	d.c. of column 1-3 by implication introduction.

A counter-example to this principle of double obligation might seem to be a situation where some act or state of affairs is legally obligatory, but, on ethical grounds, we feel that it is wrong for that act or state of affairs to be legally obligatory. Thus we would have: p is (legally) obligatory, but it is not (ethically) obligatory that p is (legally) obligatory. This is not a genuine counter-example, however, because we are really employing *two* concepts of obligation, a legal one and an ethical one. The operator 'O' could stand for legal obligation or for ethical obligation but not for both at once. It might be the case that $O_L p$ is true even if $O_E O_L p$ is not true, where ' O_L ' stands for legal obligation and ' O_E ' stands for ethical obligation. But this does not refute the view that if $O_L p$ is true so is $O_L O_L p$, or the view that if $O_E p$ is true so is $O_E O_E p$. Similarly, we might employ different kinds of obligation operators corresponding to constitutional law on the one hand, and to the laws of

various states of the United States on the other hand, and thus attempt to construct a counter-example, but the same difficulty would arise.

Furthermore, it would seem odd to assert that it is ethically obligatory not to commit murder, and still deny that it *should* be ethically obligatory not to commit murder, where the "should" itself is to be understood as expressing ethical obligation. Equally, it would seem odd to assert that it *should* be ethically obligatory not to commit murder, and still deny that it is ethically obligatory not to commit murder.

The position advocated here, indeed, is that propositions of the form Op or Pp , if true at all, *should* be true, and if they should be true, *are* true, so that we have:

$$\begin{aligned} &Op \equiv OOp, \\ \text{and} \quad &Pp \equiv OPp. \end{aligned}$$

Similarly, propositions of the form Op or Pp , if true are permitted to be true, and if permitted to be true, are true, so that we have:

$$\begin{aligned} &Op \equiv POp, \\ \text{and} \quad &Pp \equiv PPp. \end{aligned}$$

(In the above equivalences we are of course treating ' $\phi \equiv \psi$ ' as an abbreviation for ' $(\phi \supset \psi) \& (\psi \supset \phi)$ '.)

A proposition will be said to be *corrigible* if it is false but should be true, or if it is true but should be false; otherwise it will be said to be *incorrigible*. The four equivalences stated above may be said to express the principle of *deontic incorrigibility*, since they express the fact that deontic propositions (propositions of the form Op or Pp) are incorrigible. Of the four deontic systems that we have considered, DS₅ is the only one in which all four of the above equivalences are provable, and so it is the only one in which the principle of deontic incorrigibility holds. This seems to be a point in favor of assuming that DS₅ gives a more nearly correct theory of obligation than do the others.

The above equivalences, incidentally, are analogous to the following valid equivalences for quantifiers:

$$\begin{aligned} (x)A &\equiv (x)(x)A, \\ (\exists x)A &\equiv (x)(\exists x)A, \\ (x)A &\equiv (\exists x)(x)A, \\ (\exists x)A &\equiv (\exists x)(\exists x)A. \end{aligned}$$

Just as a quantifier with variable ' x ' is in effect a vacuous operator when operating on an expression in which ' x ' does not occur free, so also the opera-

tors ' O ' and ' P ', when operating on expressions of the form ' Op ' or ' Pp ', are in effect vacuous operators.

The following distributive laws are provable in the deontic systems DM, DB, DS₄, and DS₅:

$$\begin{aligned} O(p \& q) &\equiv (Op \& Oq), \\ P(p \vee q) &\equiv (Pp \vee Pq). \end{aligned}$$

The above equivalences are analogous to the following equivalences for quantifiers:

$$\begin{aligned} (x)(A \& B) &\equiv ((x)A \& (x)B), \\ (\exists x)(A \vee B) &\equiv ((\exists x)A \vee (\exists x)B). \end{aligned}$$

From the standpoint of the principle of deontic incorrigibility, we would expect the following distributive laws also to hold in DS₅, and indeed they do. Here we are letting ' d ' stand for either ' Op ' or ' Pp ', whichever we wish.

$$\begin{aligned} O(d \& q) &\equiv (d \& Oq), \\ P(d \& q) &\equiv (d \& Pq), \\ O(d \vee q) &\equiv (d \vee Oq), \\ P(d \vee q) &\equiv (d \vee Pq), \\ O(d \supset q) &\equiv (d \supset Oq), \\ P(d \supset q) &\equiv (d \supset Pq), \\ O(q \supset d) &\equiv (Pq \supset d), \\ P(q \supset d) &\equiv (Oq \supset d). \end{aligned}$$

The analogues of these last eight equivalences also hold for quantifiers. The last one, for example, corresponds to the following equivalence, where ' x ' is not free in the formula for which ' A ' stands:

$$(\exists x)(B \supset A) \equiv ((x)B \supset A).$$

Notice that the fifth of the above eight equivalences gives the result,

$$O(Op \supset p) \equiv (Op \supset Op),$$

when ' q ' has been replaced by ' p ' and when ' d ' is regarded as an abbreviation for ' Op '. This perhaps indicates why the proposition $O(Op \supset p)$ is provable in deontic systems, such as DS₅, which conform to the principle of deontic incorrigibility. We have seen that this same proposition is also provable in DB, although the latter system does not fully conform to the principle of deontic incorrigibility.

It is important to observe that the systems DM, DB, DS₄, and DS₅ all have the property that if ' ϕ ' is provable so are ' $O\phi$ ' and ' $P\phi$ '. Consequently in all these systems we have such theorems as ' $O(p \vee \sim p)$ ' and ' $P(p \vee \sim p)$ ' and also ' $\sim P(p \& \sim p)$ ' and ' $\sim O(p \& \sim p)$ '. To explain

this situation we need merely extend the principle of deontic incorrigibility to cover not only deontic propositions but also propositions which are logically true or logically false. In other words, logically true and logically false propositions are to be viewed as deontically incorrigible, so that the operators 'O' and 'P', when operating on sentences that express logically true or logically false propositions, act as vacuous operators.

The system DS₅, which from the viewpoint advocated here is the most satisfactory of the deontic systems we have considered, can be further extended by broadening its class *S* of formal sentences and then adding the usual subordinate-proof type of rules for quantifiers and identity and for necessity and possibility. We thus get a combined alethic-deontic system that employs the operators ' \Box ' and ' \Diamond ' as well as the operators 'O' and 'P', and that also makes provision for the universal quantifier and the existence quantifier and for

identity. It would seem natural in such a system to treat ' $\Box\phi$ ' and ' $\sim\Box\phi$ ' as being *obligative* reiterates of themselves, so that they could be reiterated into obligative proofs as well as into strict proofs (supposing we have an S5-like theory of necessity together with a DS₅-like theory of obligation). This would amount to treating all propositions of the form $\Box p$ or $\sim\Box p$ as deontically incorrigible, and would therefore constitute a further extension of the principle of deontic incorrigibility. On the other hand, it would seem to be unreasonable to treat ' $O\phi$ ' and ' $\sim O\phi$ ' as being *strict* reiterates of themselves, since this would give the result that whatever is obligatory is so by (logical) necessity.

A further expansion of such a system could be in the direction of making provision for dealing with properties, classes, and relations, so that various legal or ethical relations among persons could be represented.

Yale University

IV. THINGS AND DESCRIPTIONS

HERBERT HOCHBERG

IN *An Inquiry into Meaning and Truth*, Russell proposed to consider a phenomenal thing like a square white patch as a collection of qualities. He did so to avoid an alternative view which would hold that the patch consisted of a bare particular (substratum) related to qualities (universals) by the special relation (ontological tie) of exemplification. To Russell his alternative eliminated an "unknowable" that had bothered philosophers since, if not before, Aristotle's introduction of prime matter.¹ Russell recognized that a problem arises in trying to distinguish adequately one collection of qualities from another—two white squares, for example. In short, he faces the problem of individuation. This we shall take up later, for I hope to show that certain arguments purporting to establish that substrata must be recognized in order to deal adequately with the problem of individuation are unsound.² Before doing so we shall discuss some related questions that arise in the analysis of the idea of *one thing*.

Consider a white square that has only the additional property designated by " P^1 ," with " W^1 " and " S^1 " standing for "white" and "square." If "Socrates" were the name of the white square, then, on Russell's view, that term may be considered as an abbreviation for *either* a definite description of a second order class or property specifying that only P^1 , W^1 , and S^1 are members of it *or* for the set sign " $\{W^1, S^1, P^1\}$." [Let (α) stand for the description and (β) for the set sign.] Thus Socrates, like all individuals, becomes a class or property of properties.³ Such a proposal involves several difficulties. To say that Socrates is white would be to assert either " $W^1 \epsilon (\alpha)$," or " $W^1 \epsilon (\beta)$."⁴ But both statements are analytic truths. To put it

loosely, classes being what they are, the assertion that there is one and only one class having as its only members W^1 , S^1 , and P^1 and that W^1 is a member of that class is analytic, as is the assertion that white is a member of a class defined by enumeration to include white. The same would hold for assertions ascribing S^1 and P^1 to Socrates. Since the sentences asserting the existence of such a class, using either (α) or (β), are also analytic truths, that there is such a thing as Socrates also becomes an analytic truth. There are even stranger consequences. Assume that there were no white circles and no black squares, but that there was a black circle indicated either by a description analogous to (α) or by the set sign " $\{B^1, C^1, P^1\}$ " where " B^1 " and " C^1 " stand for "black" and "circle." Having such terms and properties one could then construct (descriptions or) class signs like " $\{B^1, S^1, P^1\}$ " and " $\{W^1, C^1, R^1\}$." Again, classes being what they are, statements ascribing existence to such classes would also be analytically true. To put it paradoxically, on the view that reduces things to classes of qualities, nonexistent things become necessary existents. The problem lies in the incompatibility of the logical properties of classes with the analyzing of things into classes of qualities. To avoid all this one would have to have some way of distinguishing between classes that *were* things and classes that simply *were* classes. A special property, say *existence*, at the level of properties like W^1 would not do. Instead one might suggest the introduction of a higher level property that *existent* classes would exemplify. That some such thing must be done points up the peculiarities of the position. This point also serves to contrast sharply Russell's view, which turns

¹ Bertrand Russell, *An Inquiry into Meaning and Truth* (London, Allen and Unwin, 1956), pp. 97-99.

² Arguments for particulars occur in G. E. Moore, "Identity," *Proceedings of the Aristotelian Society*, vol. 1 (1901) and Bertrand Russell, "On the Relations of Universals and Particulars," reprinted in *Logic and Knowledge*, ed. R. C. Marsh (Allen and Unwin, 1956). Moore's particulars are quality instances and Russell's may also be, but the arguments are essentially the same as those for bare particulars that later are suggested throughout G. Bergmann, *Logic and Reality* (Madison, University of Wisconsin Press, 1964), and in essays by E. Allaire and R. Grossmann in *Essays in Ontology* (The Hague, M. Nijhoff, 1963).

³ The description is $(\exists F^2)[(W^1 \epsilon F^2 \& S^1 \epsilon F^2 \& P^1 \epsilon F^2) \& (f^1)[(f^2 \epsilon F^2) \equiv ((f^1 = W^1) \vee (f^1 = S^1) \vee (f^1 = P^1))]]$. For purposes of this paper " F^2 " can be read as either a second level class or property sign and ' ϵ ' as either the class membership sign or the predicative "is."

⁴ Sometimes ' (α) ' and ' (β) ' will be used to refer and sometimes to abbreviate the description and set sign.

the sentence "Socrates is white" into a statement of class membership, with the view that predication in language reflects an ontological tie between elements of a fact or of a complex thing. Class membership as a linguistic device reflects no such tie, nor is there any on Russell's view.⁵ This is precisely what leads to the problems we just considered. To avoid such problems some device must be introduced which connects the members of *some* classes of qualities into individual entities. But then this connection will furnish the ontological tie, not class membership.⁶

One who holds that exemplification is a relation between a bare particular and a universal property might also claim that so viewed exemplification is something he comprehends in that he is acquainted with such an ontological connection. Class membership, however, is a logical relation he understands only in terms of predication and not in terms of direct acquaintance. Such a claim introduces a principle of acquaintance as a crucial theme in one's metaphysics. The issues surrounding such a principle will not concern us here.⁷ A proponent of exemplification and bare substrata might also contend that the ontological tie must connect or relate entities that are *independent* of the relation. This is not so for class membership, where a class is specified in terms of its members. In part this reflects the concern over the analyticity of sentences like "Socrates is white" on Russell's view; but it also reflects other concerns that we shall take up later. A consequence of statements like "Socrates is white" and "Socrates is square" being analytic is that all statements truly ascribing properties to Socrates are logically equivalent and, in that sense, say the same thing. Furthermore, if, contrary to the simplified case we are considering, Socrates was discovered to have additional properties, then we could not, on Russell's view, truly predicate them of him unless we altered our analysis of Socrates. One can only truly ascribe properties to Socrates that

are included in the specification of the class which Socrates is. One might then conclude that to say anything about Socrates is not only redundant, but involves knowing everything about the patch. If one adds relational properties to the class then one is on the road to internal relations and Bradley's absolute. This holistic theme will also occupy us later in another context.

Quine has proposed the replacement of proper names by definite descriptions. He has some explicit motives for doing this. One that is not so explicit may be the same as Russell's. If we use a "regular" description instead of (a) to define "Socrates" we would have (γ), " $(\lambda x)[W^1x \& S^1x \& P^1x]$." If one uses (γ) instead of the proper name "Socrates" one might feel that he is not required to recognize an entity that the proper name designates which is distinct from, independent of, and, in addition to the properties specified in the description. For, following Russell, the meaning of a description is specified by the predicates in the description; a description is not a "denoting" sign. The description can refer to something indirectly without, like a name, being connected to that thing.⁸ A name to be used independently of any description of what it names must be connected to something directly and not by means of other terms. The thing named and the connection of the name to it provide the ground for a sentence in which the name occurs, being about the thing named or referring to some fact. A description may be said to be about something in a different way. It is connected indirectly through the specified properties. This may lead one to hold that the use of a description, when there is a thing fulfilling it, reflects a consideration or analysis of that thing in terms of its properties. A name, not making use of any properties of a thing, lends itself to the idea that it refers to something about the thing other than the properties of it—the substratum or bare particular. The use of descriptions, as opposed to names, would go

⁵ Class membership is not significant for ontological questions. Part of what is involved we have just seen. That Russell's view can be stated in terms of properties rather than classes does not affect this. With "Socrates" as a second order property defined as in n. 2, predication would not reflect any ontological tie, since all ascriptions of properties to Socrates would become disjunctive identity statements that are either analytic truths or falsehoods. Where a predication is so transformed one may say no ontological connection is involved as no connection among entities is required for the statement to be true.

⁶ A referee has suggested the use of a second order property of "co-exemplification" that the first order properties exemplify when there is an individual thing. This, however, would mean that Socrates was no longer a class of qualities, but qualities in a certain relation. Such a relation would then furnish the ontological tie and hence turn the view into one we consider later. This is disguised since such a view would also require an exemplification tie so that the qualities could exemplify the relation of co-exemplification. Moreover, predication would no longer be trivial as in n. 4, unless, of course, one defined "co-exemplification" by extension.

⁷ For a claim of acquaintance with substrata and the ontological tie see G. Bergmann, *op. cit.*, pp. 47-48.

⁸ Russell, in keeping with his formula that a description is an incomplete symbol, would probably not accept this formulation.

along with a view that considered an individual to be composed solely of universals or properties in combination. The description (γ) does not reflect the turning of Socrates into a class of properties, but it may be thought to reflect his analysis into a *composite* of qualities. On the bare particular analysis, Socrates would be a bare particular tied by exemplification to universals white, square, and P^1 . On the class analysis the white patch is simply a class of qualities. A third analysis is to consider the white patch as a composite of qualities in a special structural connection or tie that would correspond to exemplification on the bare particular analysis. On this alternative the ontological tie would hold only between universals and not between universals and a further kind of thing, a substratum. The analysis of the term "Socrates" by (γ) may be taken to reflect this view. Such a view recognizes particulars, but not as either simples or substrata. Since this alternative recognizes an ontological tie that connects qualities into things, one may hold that it does not get rid of particulars in the way that Russell did. On Russell's view the lack of an ontological tie may be considered to reduce individuals to their constituent universals in a way that the present alternative does not do. [It is perhaps relevant to recall that Russell considered classes to be "logical fictions."] However, the present alternative does get rid of *bare* individuals or substrata. If defensible, it thus reaches Russell's goal.

Before pursuing the above point let us compare the use of (γ) with the use of a proper name like "Socrates." Consider three sentences, using (γ), to assert "Socrates is white," "Socrates is square," and "Socrates is P^1 ." None are analytic truths. Nor would a sentence asserting that the description (γ) is fulfilled be an analytic truth. This points to a radical difference between (γ) and (α). Furthermore, the problem about the non-existent black square does not arise in the case of (γ). Yet the three statements asserting that Socrates is white, square and P^1 are logically equivalent to each other and to the assertion that the description (γ) is fulfilled. In this sense all these statements, using (γ), say the same thing.⁹ But this is neither surprising nor detrimental. To name Socrates is only to indicate him. Where the term "Socrates" is a name, the sentences "Socrates is white" and "Socrates is square" each state that

what is indicated by that name has a certain, and different, property. A description does not *merely* indicate. It indicates by means of properties purporting uniquely to determine an object. Hence, to ascribe such properties to the object, indicated by means of them, is to do something different from indicating by means of a name. Moreover, there is an analogous, though not explicit, feature in the use of names. When seeing a white square and naming it or referring to it by a purely indicating sign like "this" or a proper name, we do so in virtue of something we notice about it. That is, one does not come across a substratum or particular apart from properties. One confronts it, if at all, exemplifying properties. In applying a name to a bare particular or substratum proponents of such things think of distinguishing them from the properties they exemplify. This may be looked upon as just another way of saying that the sign used as a proper name has only an indicating function. Hence, a bare particular becomes a hypostatization of this function of a sign. The use of descriptions makes explicit the fact that where something is indicated properties of that thing play a role. Proposing the use of definite descriptions in place of proper names may thus reflect the rejection of bare particulars as elements of one's ontology. This may be gotten at, alternatively, by holding that proper names name complexes of qualities, yet are simple, primitive signs. One who argues in this way rejects the notion that language must picture objects and hence that simple signs cannot designate complex objects. The proponent of descriptions, as opposed to names, as indicators of complexes of qualities might then be thought to accept, *to a degree*, such a "picture principle" of language while rejecting bare particulars as entities. Since his individual objects are composites of universals, the signs corresponding to such individuals must be composites of signs which refer to the universals involved. The composite sign is linked to the complex object it corresponds to since the signs it is composed of refer to the entities the object is composed of. Furthermore, while descriptions, as complex signs, would correspond to complexes of qualities as complex entities, they would do so only to a certain degree. That is, a description need not contain predicates indicating all the qualities of the indicated object. An object described by (γ) on the present alternative could have

⁹ The sense in which these formulations say the same thing is like the sense in which " $2 + 2 = 4$ " and " $7 + 3 = 10$ " say the same thing. Also with respect to intentional contexts none of these statements say the same thing as any other one.

further properties. Thus another problem that arose on Russell's view does not arise. Suppose, a bit more realistically than our simplifying assumption that Socrates has properties in addition to W^1 , S^1 , and P^1 , but that predicates referring to these additional properties do not enter into his description in (γ) . To ascribe any such additional properties to him will not in the least be to say "the same thing" as ascribing properties included in the description. Nor will any two such ascriptions say the same thing as each other. Thus a further difference to Russell's view is involved. Moreover, one must not be misled by the following argument. Suppose W^1 , S^1 , and P^1 suffice to individuate Socrates, then if in addition he has R^1 and Q^1 , the set of properties W^1 , S^1 , P^1 , R^1 , and Q^1 will also individuate him. Let (γ^1) stand for a description constructed from this latter set of predicates. Then the identity " $(\gamma) = (\gamma^1)$ " will hold. We could then replace the definition of the term "Socrates" by the more extensive descriptive phrase. One could do this for all the descriptive properties of Socrates and hence turn any statement ascribing one of these to him into a statement logically equivalent to any other statement ascribing a different property to him. The point to be made in reply is that the above identity statement is synthetic.

Those who advocate a bare particular analysis and the naming of such things still, generally, hold that one indicates past objects or things one is not now acquainted with, and hence not capable of being simply and directly indicated, by definite descriptions. Hence, to say about such things that they have or had certain qualities involves the same seeming redundancies. On a bare particular analysis one avoids these seeming redundancies at the price of introducing two kinds of entities. First, there is the mysterious bare particular which is named. Second, this entity combines, by predication, with a universal to form a fact which determines the truth of a sentence asserting that the thing has the quality. Such a true sentence may even be thought to refer to this further thing, the fact. On the alternative analysis, since the white patch is considered a combination of qualities no further entity is involved. A true sentence ascribing a quality specifies a constituent of the complex indicated by the descriptive phrase. It does not assert that something is related to what is indicated and hence refer to a third, complex, entity com-

prising the two relata in a relation of predication. Holding that Socrates is a composite of qualities one might say that it is the complex thing that makes the sentence true.

All this brings us to the question of individuation, for it is on this issue that the bare particular analyst bases his case. Supposedly, alternative views cannot account for individuation and difference.

Quine sought to solve the problem of individuation by introducing peculiar individuating properties and basing descriptions on predicates referring to such properties—"Pegasizing." He made one mistake in holding that such primitive properties could be introduced for unfulfilled descriptions. Thus he had signs whose only function was an indicating one with nothing being indicated. But we are not here concerned with questions about references to non-existent "things." Talking about existent things having a unique property, which serves solely to individuate them, simply puts proper names (and bare particulars) at the predicate level. Such properties of individuation are certainly as puzzling and mysterious as bare substrata, and, like these latter, are hypostatizations, in a more devious way, of the simple indicating function that some terms may have.

Russell tried another approach. Consider two white patches, Socrates and Plato. Assume that they are alike in all non-relational descriptive properties and that Socrates is to the left of Plato. To distinguish them qualitatively Russell assigned to each the property of being at a place in the visual field. Some would object to this on the grounds (a) that they are not acquainted with such properties of things, and (b) that space is relational, while on Russell's proposal it is not.¹⁰ But is Russell's suggestion so outlandish? Given a succession of phenomena cannot one recognize that a patch is in the same part of the visual field as a previous one, just as he recognizes that it is the same color as the previous one? However, given a succession of exactly similar visual fields one might also have to have recourse to similar properties of a temporal kind. Perhaps such properties cannot stand up to philosophical probing. But this question can be waived here, for one has had recourse to such properties due to the acceptance of certain arguments that forbid the use of relational properties for purposes of individuation. These arguments, I wish to show, are not cogent and, consequently, one may employ

¹⁰ G. Bergmann, "Russell on Particulars," reprinted in *The Metaphysics of Logical Positivism* (New York; Longmans, Green & Co., 1954).

relational predicates in descriptive phrases to indicate Socrates and Plato.

Suppose that in order to distinguish the descriptions of Socrates and Plato, where Socrates and Plato are two exactly similar patches with Socrates being to the left of Plato, one proposes to include the description of Socrates the predicate "being to the left of Plato." This would immediately be seen to be inadequate since the definition of such a predicate would have to include the descriptive phrase indicating Plato. This in turn would have to include the predicate "being to the right of Socrates" or it would not have been uniquely specified as distinct from Socrates, for, recall, all its non-relational properties are shared with Socrates. But this problematic situation is easily changed. Suppose one introduces the property L , with " R " standing for "right of," by

$$Lx = \text{df } (\exists y) (Ryx \ \& \ Wy \ \& \ Sy)$$

and suppose further that no white patch is to the right of Plato. For a thing to have L is to have a white square to its right. To the use of such a property the proponent of bare particulars retorts that to use a relation we must already have terms standing in that relation, and hence to use a relational property as a constituent property in forming a definite description is illegitimate. This argument is confused on two counts.

First, it confuses "being about" a white patch in the sense of saying that something is white with formulating a descriptive phrase to indicate or be about the thing. Thus in formulating the descriptive phrase to indicate Socrates one refers to Plato only in that one talks about something being to the right of Socrates. It is not a question of using the descriptive phrase for Plato in the construction of a relational property to individuate Socrates and of using the descriptive phrase for Socrates in the construction of such a property for Plato. Consequently no circularity is involved.

Second, one might, to use a cryptic and loose phrase, say that *logical* and *temporal* priority are confused. Perhaps the point can be clarified in the following way. One must indeed notice Socrates as distinguished from Plato. One can also refer to these different things by different names. This does not mean, having done this, that one cannot consider an analysis of the things in terms of properties and of the names in terms of descriptions composed of predicates referring to those properties. Two patches are noticed to be different and in a spatial relation. To notice them in a

spatial relation does involve that instance of the relation depending on there being things related. But this does not mean that the things related do not also depend on that instance of the relation. Without there being some spatial relation between them "they" would be one and not two. Nor does it mean that such mutual dependence prohibits the use of the relation in an analysis of the things. One might think otherwise if he confuses a relation with an instance of it and thinks that " L " must be defined in terms of "being to the right of Plato" instead of in terms of " R ". Also, if one holds that exemplification is the ontological tie or relation and that such a tie must hold between independent relata, he might hold that all relations require ontologically independent relata. It is understandable why one would hold that exemplification requires independent and simple relata. Recall the sentence "Socrates is white." If the subject term is held to refer to a composite entity that contains whiteness, and thus *depends* on the universal, one might feel that predicating "white" of Socrates is redundant, or empty or analytic. This we discussed earlier. Furthermore, one who adheres to a picture principle of language would naturally believe that predication in language ought to reflect the ontological tie between the simple elements of an ontology. On the view adhering to substrata, universals, and exemplification it seems so. On the alternative view it apparently does not since the description (γ) refers to a complex in which both the tie and the universal are constituents. The ontological tie on this view combines simple universals into things; it does not combine a substratum and a universal into a fact. This difference is behind the fear that predication without substrata is empty. Be that as it may, that relations are not presented without relata does not mean that one cannot analyze particulars in terms of qualities and relations. Ontology is not phenomenology. Thus even if one may notice something without being aware of what relations it stands in or properties it has that fact has no bearing on the issue. To the bare particular analyst it might. Thus he may argue for numerical difference as distinct from conceptual difference by holding that he can apprehend our two patches as different without noticing how they differ. One might then think that if he notices *simply* that two things differ, they must differ in *simples*. Since they do not differ, in our example, in simple non-relational properties, and, since relational proper-

ties "involve" terms (and hence are thought incapable of grounding simple difference), they must differ in simple bare particulars. From what one simply notices one is thus led to a kind of ontological simple. Since, via a principle of acquaintance, one convinces oneself that a bare particular is an object of acquaintance, the circle is closed and the knot is tied. The bare particular is the ontological representative of apprehended difference and the ground of numerical difference. We may protest that all this is far too simple. For some might feel that they just don't know what it is to apprehend two particular things as different without apprehending them to be different in some way. But that aside, the above argument confuses what I apprehend with the analysis of what is apprehended. To put the matter slightly differently we may say that without there being a difference in property (including relations) between Socrates and Plato there would be no apprehension of simple difference. That this is so points up that what "makes" them different is one question. Whether they may be apprehended as simply different is another. It is relevant to notice that for the bare particular analyst there is the possibility of having Socrates and Plato be different while not differing in any properties or relations, for bare particulars may be simply different, and, I take it, could be apprehended as such. That this does not happen is, from the perspective of the bare particular analyst, a fact about our world. Looked at from another perspective it reveals an inherent absurdity in the bare particular analysis. In any case, as I am said to notice something being simply different from something else, it is then thought that this difference must not, ultimately, be grounded or analyzed in terms of properties or relations. In addition to the problematic nature of this assumption, there is the sheer question-begging element involved in the notion of "something." If "something" is taken as a bare particular then, of course, from the very role such a thing plays it is simply different from another such thing. But if the something is taken as the patch from which we start our ontological analysis, then the matter is not closed. We may still consider it to be a composite of qualities. If at this point one argues that it cannot be analyzed as a composite of qualities, because it is just seen as a distinct thing, not as a composite, I hardly know what to say, except to repeat that ontology is not phenomenology. Of course Plato and Socrates are apprehended as distinct things.

This is where analysis begins. But this does not mean that the bare particular analysis is the correct one. Some may think that it does—that to say that bare particulars account for the difference of Plato and Socrates is simply to say that the latter are different. This explains why one may also convince himself that he is acquainted with bare particulars. It also points up that in attempting to defend bare particulars as entities and as objects of acquaintance ontology may be reduced to triviality.

We may then reject the arguments we have considered in favor of bare particulars and allow a description which employs a property like L to indicate Socrates. For objects, one need not recognize numerical difference as distinct from conceptual difference. For what makes Socrates differ from Plato may be held to be a quality occurring in his description. In this context, we might note something about numerical difference. Suppose in order to reflect that relation we introduced into a language schema a predicate " D ," as a primitive predicate distinct from " \neq ," the latter being defined in the Russell-Leibniz fashion as " $\sim(F)(Fx \equiv Fy)$." Whatever else the property D would involve it would be such that (δ) " $(x)(y)[x \neq y \supset Dxy]$ " would be true. Otherwise one would have conceptually different things being numerically the same. But it is difficult to consider what that would mean. I cannot. This is part of the peculiarity about " D ." It is, supposedly, a primitive predicate, yet (δ) is hardly an empirical generalization. Hence (δ) becomes either a synthetic *a priori* truth, a partial or implicit definition, an addition to the analytic statements of the language, or what have you. In short it becomes some form of necessary truth. Consider two things that are numerically different, a_1 and a_2 . a_1 will not be numerically different from itself, hence there will be a context involving the property D that distinguishes a_1 from a_2 . That something does not differ numerically from itself will, I take it, be a "necessary" truth like (δ) . Hence, it will "follow" from two things being numerically different that they are conceptually different. For, to put it paradoxically, D is a concept among concepts. Therefore, (ϵ) " $(x)(y)[Dxy \supset x \neq y]$ " will also be a kind of necessary truth. [Just as "follow," above, reflects a kind of "inference."] Thus " D " and " \neq " are, in some sense, "logically" equivalent notions. All this points up both the redundancy and peculiarity of " D ," as a primitive predicate, and why

numerical difference is represented in the language schema, not by "*D*," but by the occurrence of different signs for different things. Numerical difference, like conceptual difference, is said to be "logical," but in a different sense. The relation expressed by " \neq " is logical in that it is defined in terms of logical signs—connectives, variables, and quantifiers. Numerical difference is logical by analogy to predication. The latter is reflected in a language schema by a syntactical device, say juxtaposition and the type distinction, to show that what is reflected is a structural or ontological tie rather than an ordinary relation and to acknowledge and avoid the puzzles associated with Bradley. But the problems about "*D*" are not on a par with Bradley's puzzles about predication. Any ontology must acknowledge some *connection* or *tie*, corresponding to predication, and distinguish it *in kind*, to avoid Bradley's problems, from what it connects or ties. Numerical difference is another matter. To put it cryptically one doesn't need either it or bare particulars in the way in which one needs a tie like predication. Hence the peculiarities about "*D*" reflect, not a general or logical feature of ontology, but puzzles arising from a particular solution to the problems of ontology. This may be ignored by assimilating numerical difference to predication as a basic logical or structural or categorial feature of reality.

To the view that Socrates and Plato are composites of qualities, including relations, it may be objected that this forces one to acknowledge that a thing changes when what it is related to changes. In short one is involved with *internal relations*. Thus if *L* is included in the set of properties that constitute Socrates, the replacement of Plato by a red square would mean that Socrates was no longer what he was. Yet, if one is talking about phenomenal things he may well, for a variety of reasons, hold that such things neither change nor persist through change. That Socrates does not persist through a change—the disappearance of Plato—would be perfectly in keeping with this contention. Even if one does not restrict himself to phenomenal entities, but rejects continuants, the point would be the same. Actually, the bare particular analyst faces the same problem in his own way. Given that Socrates and Plato are bare particulars and that the latter disappears then to hold that Socrates persists through time forces him (1) to acknowledge that qualities, relational or otherwise, are attributed *at a time* rather than simply predicated, (2) to hold that the persistence

of Socrates is not to be taken literally since another particular exactly like Socrates is what is to the left of the red square after the change, or (3) to divide, but not reduce, Socrates into parts coincident with the period before and after the change; this would involve there being three bare particulars—Socrates and his temporal parts—related in certain ways. (1) and (3) introduce continuants. (2), in rejecting continuants through relational change, reflects the point about internal relations. Here we are not concerned to explore the issues surrounding (1), (2), and (3) on a bare particular analysis. The point simply is that the same sorts of problems face such an analyst. Further, certainly a variant of (1) is possible on the analysis of Socrates into a composite of qualities if one makes use of predication *at a time*. Unlike (3), however, a composite, as opposed to a bare particular, is not literally the same composite through a change of a constituent. But the peculiarity of (3) reveals the problem of acknowledging continuants on a bare particular analysis. The issue is then whether there are continuants through relational change. If one is convinced that there are no continuants at all, then the analysis of Socrates into a composite of qualities including relations would cause no anxiety. If one is convinced that there are continuants, then perhaps he will argue for bare particulars on such a ground. But this is to base his contention on arguments differing from those we are rejecting in this paper.

A proponent of bare particulars may raise two further objections to the view we are considering. First, he might hold that the use of descriptions begs the question since (1) zero level variables are employed, and (2) the connection of predication is used. But, while zero level variables are used, composites of qualities not bare particulars are, as some say, the "values of the variables." And, the linguistic relation of predication need not reflect the ontological connection between a bare particular and a quality. What it *reflects* is the ontological connection between qualities in a composite. Further, it is used to specify a member of that composite in a true sentence about it. Second, he might point out that to hold that things are composites of qualities is to turn predication between zero and first level signs into something different from predication between first and second level predicates. To say that green is a color is not to say that green, a simple, is a composite containing color as a constituent. Yet one might

wonder if exemplification between a bare particular and a universal is the same thing as exemplification between two universals. It is the same in the sense that it holds between simples in both cases and that in both cases the simples are thought to combine into facts. It is different in that the simples are of logically different types. Moreover, only the substrata and first level universals are held to be constituents of the things we started out to analyze. This crucial difference indicates that exemplification, on the substrata analysis, does not function uniformly. The question invites detailed analysis. The point here is that it is not obvious that the bare particular analysis requires fewer ontological ties. If that analysis does require fewer such ties then this would be a point in favor of it. But this is a different argument for bare particulars than those we are rejecting.

A definite description need not specify all the properties of what is described, only those sufficing to individuate. One might feel that it should mention all, eventually, since then the descriptive term will reflect, in language, what sort of entity is being described—its structure so to speak. For, it will show that it is a composite and what it is composed of, just as, on a bare particular analysis, a proper name referring to a bare particular reflects, in language, that it is and what it is—a simple particular. Thinking of the predicates in a description as furnishing a meaning for the descriptive phrase, one then thinks the addition of predicates to a description reflects the growth of meaning. Including relations of all kinds can lead one to hold that to know what a thing is, i.e., to know what the term indicating it means, is to know everything. Hence including relational properties in the descriptions of things leads one to Bradley's holistic Absolute. Also, one is led to the complete redundancy of all statements about something when using such a comprehensive description. Alternately taking a bare particular as the meaning-referent of a proper name avoids such redundancy. But one must not be misled by the way this matter is put. A fulfilled description indicates or is connected with a thing by means of the predicates in the description. We can then distinguish two aspects of a description: first, its

indicating role and, second, its meaning in the sense in which the latter is specified by the predicates it contains. In view of the second one may be led to hold that of two descriptions of the same thing the one that specifies more properties provides more meaning. But all this could mean is that since one answers the question "What is this?" by listing properties a more complete description provides a fuller answer to that question. If we keep separate the questions (a) "What is it?" in the sense of what are its non-relational properties; (b) "What is it?" in the sense of all the things one can say about it; (c) "What is it?" in the sense of what properties need be specified in order to individuate it; (d) "What is it?" in the sense in which the answer is simply "a composite of qualities"; (e) "What is it?" in the sense in which one asks "What is being indicated?" and in which the answer might be simply "this"; then no harm is done. Mixing different senses of this "question" can lead one, on the one hand, to invent bare particulars, and, on the other, to Bradley's absolute. That one could indicate a complex of qualities by the term "this" or a proper name does not mean that there need be an additional simple entity which is what is really named. To think so is to accept uncritically the principle that a simple term must indicate a simple thing. One who feels forced to indicate a composite of qualities by a complex sign such as a definite description also, in a way, accepts this principle. But such a principle is not bound to the view that things are complexes of qualities. One may on rejecting this principle indicate such complexes of qualities by proper names—signs with a purely indicating function. This is another matter, though we may note here that it is an alternative that could avoid the consideration of relational properties in the complex that constitutes the thing.¹¹ (Two complexes of non-relational qualities would just be different as bare particulars are just different.) One might suggest that both the bare particular analyst and the holist accept the principle that language must picture reality in an extreme form. Just as one invents a simple thing to correspond to a simple term, the other holds that the corres-

¹¹ What I have called "Russell's view" is not, literally, what he proposed, since he treats the sign "Socrates" as an undefined sign (a proper name) to avoid turning sentences like "Socrates is white" into analytic truths. But as he analyzed objects into classes of coexistent qualities and did not recognize the need for (or role of) an ontological tie, his use of names for classes is specious. To argue this point is beyond the scope of this paper, though part of what is involved is touched on in notes 5 and 6. The use of proper names for objects considered as complexes, but not classes, of qualities is discussed in "Things and Qualities" in the proceedings of the Oberlin philosophical meetings for 1964.

ponding term must reflect all the complexity of the complex thing. Hence, in order for a description to indicate a composite of qualities it must indicate all of them, including relations. As our knowledge of a thing grows *what the thing is* is revealed by the term that indicates it—by its *meaning* or *definition*, which is specified in terms of

its constituent predicates. We then arrive at absolute idealism and the notion that all statements about a thing are redundant or analytic. All that is then left to be said about “what it is” is that it is what it is—“Reality is reality.” This illusion and that of bare particulars are opposite extremes of the same kind of mistake.

Indiana University

V. ON EXPERIENCE AND THE DEVELOPMENT OF THE UNDERSTANDING

PETER UNGER

I. Empiricist philosophers usually agree on a doctrine which I doubt is clear to any of them and which, if they would examine sufficiently, most would dissent from. This is the traditional doctrine concerning the development of the understanding.

The original doctrine asserts that for a person to have fully developed concepts he must have experience. To be more specific, the development of certain particular concepts is dependent on a person's having particular sorts of experience. These concepts may be called empirical concepts and this experience the kind of experience necessary for the development of the empirical concept involved. (One may call this experience the experience of a kind of "thing" which the concept is a concept of.) Additionally, the development of concepts other than empirical concepts is dependent upon the development of particular empirical concepts. These other concepts can be said to be derived from the particular empirical concepts on whose development their development is held to be dependent.

I wish to attack this traditional doctrine and prove that, as a purely conceptual doctrine, it is false. When the formulae used to express this conceptual doctrine are interpreted so as to express a psychological generalization they may well express a truth, though I very much doubt that this is the case. But even if this were the case, it is not of particular interest to me in this essay.

More explicitly I wish to show the following *conceptual doctrines* to be false:

(1a) The full development of a certain set of concepts, called empirical concepts, requires experience, and consequently,

(1b) The full development of a certain set of concepts, called empirical concepts, requires experience and the development of any other set of concepts depends upon the learning of this first set of concepts.

Using the concept of red as an example one

may express 1a in any of the following more specific forms:

(2) A person who understands the nature of red must have¹ experience of red.

(3) A person who understands what the word "red" means must have experience of red.

(4) A person who understands what the word "red" means must have learned something; namely, what the word "red" means.

(5) A person who has the concept of red must have learned something.

Also interesting to refute will be the following more conservative counterparts of doctrines (2), (3), and (4):

(2') A person who understands the nature of red must have experience.

(3') A person who understands what the word "red" means must have experience.

(4') A person who understands what the word "red" means must have learned something.

Pursuing the matter further, I shall also refute the following yet more challenging forms of these doctrines:

(2*) A person who fully understands the nature of red must have experience.

(3*) A person who fully understands what the word "red" means must have experience.

(4*) A person who fully understands what the word "red" means must have learned something.

(5*) A person who has a fully developed concept of red must have learned something.

Again, any of these formulae so interpreted as to express "psychological generalizations" may very well be true; but even if so, this would not be of particular interest to me in this essay. I am concerned with these formulae only so interpreted as to express conceptual doctrines; I wish to show that if so interpreted, they are false.

To be still more specific about my task, and yet generalize my results still further, I exhibit the following doctrines which our method of argument will be specifically designed to *establish*:

¹ By "have" I mean "presently have or have had."

*(A) For any x , where x is something that can be fully conceptualized by a human being, it is not the case that a person who has a fully developed concept of x must have experience.

(A') For any x , where x is something that can be fully conceptualized by a human being, it is not the case that a person who has a fully developed concept of x must have learned something.

(B) For any x , where x is a word that can be fully understood by a human being, it is not the case that a person who fully understands what x means must have experience.

(B') For any x , where x is a word that can be fully understood by a human being, it is not the case that a person who fully understands what x means must have learned something.

Before I begin my arguments, I attempt to characterize my opposition, though only very generally. Here we summarize their apparent thoughts, going on to present their motivation in Section 2.

In the writings of the early Empiricist philosophers it is unclear whether the authors intend to express a general psychological truth, or whether they think that they are stating a "deeper," non-psychological truth or not-merely-psychological truth. It is a bit of philosophical patter to say that the early Empiricists were not very good at telling the difference between philosophy and psychology; or that they had not clearly formulated this difference. Later Empiricist philosophers and philosophers of an Empiricist persuasion seem to be clear that they do not mean to state empirical, psychological truths by asserting these doctrines. Rather they wish to assert conceptual truths, to make conceptual points. It is these modern Empiricists that I am most concerned to prove wrong.²

2. Why have these doctrines been adopted by so many philosophers? Perhaps the following thoughts could explain this adoption:

It is believed that experience is infused with sensible qualities which the experience is experience of. In having the experience of red I necessarily am aware that nothing but this kind of experience could show me what red was. I am immediately and necessarily aware that to understand fully the nature of red, I must have an experience of this sort.

One looks at something red and says to oneself:

"How could I possibly know what red was were it not for *this* (experience)?" When this sort of question is put, no answer is forthcoming, and so one concludes that there is no other possibility but that asserted by the Empiricist doctrines. It is difficult to consider that perhaps one is not asking the right question and should ask instead a question which is more perspicuous, and more relevant to the issue under discussion.

3. It seems obvious to common sense that a congenitally blind man cannot fully have the concept of red. (However, some may conclude that their analysis of the concept of a concept shows that congenitally blind men can have a full concept of red.) But, I do not wish to discuss whether or not blind men can have a full concept of red because (a) I think it unnecessary to the issue at hand and will prove that this is so, and (b) proving that a congenitally blind man *could* have a full concept of red would not prove as much as I wish to prove. For all of the doctrines, save (2) and (3), stand as they are even if it is shown that congenitally blind men could have a full concept of red. For it remains an open question whether the concept of red that a congenitally blind man would have would depend upon his having other experience (we may suppose non-visual experience). *Prima facie* one may maintain in the face of any proof directed solely at this particular question that the other experience that a blind man may have is essential for the development of the concept of red in him. (Furthermore, there are fairly trivial proofs that congenitally blind men could have a full concept of red, based on an adequate analysis of the concept of perception, in particular, visual perception. But I need not discuss this here.)

4. In arguing against these doctrines of Empiricism it is best not to begin by asking "How could I know this but for this?" but rather to begin by asking the question "Could there be a person such that he did have fully developed empirical concepts (etc.) but did not have any experience?" One then tries to find a case where one would say that a person had fully developed empirical concepts, and where one would not say that this person had any (previous) experience; but on the contrary would say that he did not. One asks similar, appropriate questions of cases designed to prove other of the doctrines false;

² For examples, one might note such otherwise differing philosophers as: H. H. Price in his *Thinking and Experience* (Cambridge, Harvard University Press, 1953), pp. 52, 53, 57-58; Bertrand Russell in his *Human Knowledge: Its Scope and Limits* (New York, Simon and Schuster, 1948), p. 499; and Gilbert Ryle in his *The Concept of Mind* (London, Hutchinson, 1962), p. 317.

e.g., the ones which mention learning but not experience. (In philosophic jargon: one looks for certain kinds of sufficient conditions, those in which there is an absence of what is claimed to be a necessary condition.) It would simplify the work greatly if one could find one case in which one should get the proper, sought after answer to all of the questions: questions appropriate to all of the doctrines which I am trying to disprove. There is such a case and it is not so hard to find.

5. I begin by considering the following situation: A man, whom I know has had experience of red, perhaps myself, is placed in an environment where nothing is colored red. Say, he is in a yellow room containing no articles but green ones. Then he is asked, perhaps by way of a loudspeaker system (and *via* a tape-recorded message), whether he has the concept of red, whether he understands what red is, etc. He replies, naturally enough: "Of course, I do. What are you up to anyway?" Then come many other questions about the color of fire engines and many other equally interesting subjects, which he answers with great and undoubted success. After he leaves the room the loudspeaker system "asks" him whether he had any experience of red while he was in the room—even that "of his own making." He honestly, and I may *suppose*³ truly, replies that he had no experience of red while in the room, none at all, not even "of his own making." He is then led into another room where he undergoes tests to decide whether he can see red, and whether he understands what "red" means. Of course, he passes these with flying colors. Even when asked to "conjure up an image of red," say an image as of a red triangle, he, I may *suppose* truly, reports success. (Whether or not and why conjuring up images should be considered important is immaterial. The fact is that I may *suppose* that he can do so; so I need not discuss the issue.) This is a man who undoubtedly has a fully developed concept of red. But, equally, this is a man who has prior experience of red, who has learned something, etc., as well.

Here I put forward the essential consideration: I may suppose that scientists construct another man, cell-part for cell-part, cell for cell, nervous structure for nervous structure identical to the man we have just considered, i.e., the first man. This second, constructed, man is so made that he is an

exact physical duplicate of the first man as the first man was upon entering the aforementioned yellow room. In particular, the state of the nervous system of the first man when he entered the yellow room is physically reduplicated in the construction of the second man. Additionally, I *stipulate* that the second man is placed under a wholly effective anesthetic throughout his construction.

The second man is then placed in the doorway of the yellow room. The room is in just the same condition it had been in when the first man did his rounds. And the second man is in just the "position" that the first man was in when he entered the room. The anesthetic which the second man was under is now removed, and as both the stimulus conditions and the conditions of the organism are just as they were in the case of the first man, the second man exhibits an identical pattern of behavior. In particular, his responses to the identical loudspeaker-produced questions are answers identical to those of the first man. (As everything is the same in both cases, contemporarily popular distinctions, e.g., that between actions and movements, need not be considered. Indeed, I'm inclined to think that this is probably the case with respect to a very wide range of fundamental questions in the traditionally central areas of the Philosophy of Mind and the Theory of Knowledge.)

The second man then continues through the sequence, going into the second room and, identically, passing all of the tests for determining whether someone had the concept of red, knew what red was, etc. In particular, I may *suppose* that he *truly* reports that he was successful in calling up images when asked to do so.

I should think that philosophers, including philosophers who think of themselves as Empiricists, should conclude that the second man did have a fully developed concept of red, did fully understand what the nature of red was, did fully understand what the word "red" means, etc. These facts were true of him when he was in the first room where he had no experience of red. And, prior to entering the first room, the second man had no experience whatsoever. Also, prior to entering the first room, the second man had learned nothing. No learning had taken place.

6. Someone may remark that the second man, although he did not have any experience of red

³ The word "suppose" is italicized to provide a reminder that the issues of this essay in no way involve the sceptical problems of epistemology, in particular, the celebrated problem of other minds.

while he was in the first room, did have some (other kinds of) experience then. The remarker then may think that having this other experience is essential to the truth-conditions of the statement. But surely this would be an erroneous thought. For I may suppose that after the *first* man had passed all of the tests (in the second room) he went to sleep; he had a very deep sleep perhaps aided by a powerful anesthetic. I may *suppose* that while he was asleep he was having no experience whatever. A man, whom we may call the third man, is made to be an exact duplicate of the first man as he was when sleeping. The two men, the first and third, eventually awake and, in an identical way, manifest complete understanding, etc., of what red is. Both men fully understood what red was, etc., while they were asleep; that is, it was correct to say of each of the sleeping men that he fully knew what red was, etc. Yet the third man had no previous experience and no experience contemporary with the first moment at which it would without any reasonable doubt have been correct to say of him that he fully knew what red was, etc. This because he was then fully asleep.

Nor was the experience which the third man had after he awoke essential to his understanding, etc. For we may build a fourth man just like the third man (and so the first) whom we do not allow to wake up (may the thought perish with the man). It is true of this, fourth man, that while asleep, alive and whole, he was a person who did fully understand what red was, etc. And here we have a person who never had any experience whatever, and never learned anything.

All of the main questions, it is easy to see, can be asked and answered, solely by concentrating one's attention on the case concerning the second man. The cases involving the third man and the fourth man are just mentioned to meet particular objects and clarify particular misunderstanding. My method of proof is sweeping: it puts forward a single case which disproves all of the Empiricist doctrines under discussion. Also, it establishes all of the statements: A, A', B, B'. This because our duplicate can do everything we can do although he has had no previous experience and has never learned anything.

7. I require for the correct performance of the model in my proof roughly that (I) it is possible that states of a finite physical system, e.g., a human body, reoccur, and *perhaps* that (II) each

"mental state" is correlated with a particular physical state, and if a particular "mental state" of a person is ever correlated with a particular physical state, then whenever this kind of physical state occurs in the world, this kind of mental state will be correlated with it. I need only the second requirement if it is true that there are such things as "mental states" which can (logically) be correlated with particular physical states, and that whether or not one has any of these is a factor to be considered in determining whether or not one has concepts, etc. These are requirements which I believe most contemporary philosophers would grant, especially most of those who think of themselves as Empiricists. Certainly they would not deny these requirements on conceptual grounds, save the possible exception of denying the requirement concerning "mental states" on the grounds that talk of such states is meaningless, etc. But this denial (behaviorism, materialism, etc.) is perfectly all right with me, for one needs this second requirement only to satisfy philosophers who are more dualistically inclined.

The important point for us to note is that if we *deny* either of these requirements we are *not* making a logical or conceptual remark in doing so.⁴ We are denying that as a matter of fact things are a certain way. And this is all I really need to consider; so my proof emerges completely unscathed.

8. The examples exhibited in Sections 5 and 6 clearly refute the Empiricist doctrines. Once the doctrines are clearly refuted, it is clear that they are refutable. Now, when it is clear that the doctrines are refutable, it may be tempting to exclaim that it must always have been clear that this is so. For if the doctrines could not be rebutted by counter-example, then that must be because the doctrines are analytically true. And surely these doctrines are not analytic statements. Indeed, they do not even begin to have the look of analytic statements. So it cannot be surprising that the doctrines admit of contradiction by counter-example. Indeed, why go to all the trouble of concocting examples so detailed as those of Sections 5 and 6? Why not just say that as the doctrines are not analytic, it is *possible* that a man who grew up in a dark closet could emerge therefrom and simply have a fully developed concept of red? He would *do* anything which a person with a fully developed concept of red *would do*, so why

⁴ Save in the aforementioned case of denying the second requirement on conceptual grounds. But this denial (behaviorism, materialism, etc.) is sufficient to insure the success of my proof, given the fulfillment of the first requirement.

shouldn't he have the concept and fully? Surely, any denial of or resistance to this possibility can only be due to thinking that the doctrines are analytic. But there isn't even the appearance of a reason for thinking that they are. That is all there is to the matter.

But is that all there is? If so, then how are we to interpret the Empiricist doctrines? Are they simply truths of human nature, statements of simple fact? Are they psychological generalizations? If so, then, barring our exhibited examples, what could show them to be wrong? "The man from the closet," one might say.

So let us reconsider the case of the man who grew up in a dark closet. *Ex hypothesi*, when this man emerges from his closet background, he acts in such a way that anyone who saw him but was ignorant of his previous history would say that he had fully developed concepts. But a sceptic might reply either:

(A) The person, due to some "mechanical" goings-on, acted just like a person who really understood what red was, etc.; but he really didn't understand what red was, etc., because he did not have the appropriate previous experience.

or

(B) The test situation which the subject is now facing initially provides him with the experience necessary for the development of his understanding of the nature of red, etc. And then moments later, the subject successfully passes the tests due to the experience gained in the first moments of exposure to the test situation.

The complexities of human mental life and the necessity of perceiving any situation which is to be a test situation, not only allow for, but encourage the replies (A) and (B). These sceptical replies allow us to retain the Empiricist doctrines by an interpretation of the hypothetical phenomena in terms consistent with these doctrines.

To my own mind at least, the sceptic's replies are worth philosophical consideration. For they have the quality of remarks put forward by one who is defending a position which strikes him as important and necessarily true. Indeed, some motivation which might lead to the sceptic's replies has already been indicated in Section 2, and this motivation should not be considered trivial. When I think of the various qualities in the world, it does seem that I would not fully understand their nature unless, somehow, I was *shown* what the nature was. And in order that this be *shown*,

ultimately I must be shown it by my *experience*. The doctrines *prima facie* strike me as necessarily true. If all necessary truths are analytic statements, that is necessarily true simply in virtue of the definitions or rules of some natural or formal language, this does not change the situation.

I should think that one who held an "analytic" position on the question of necessary truth would think that among the analytic statements are tautologies which are substitution instances in theorems of first order logic. Notice, then, that with such analytic statements, we are often surprised to find that they are analytic. Analyticity cannot always be immediately recognized. And in identical manner, one cannot always immediately recognize as contingent statements which are of that kind. For the expression of a contingent statement can equally involve a fair degree of logical apparatus, and then calculation will normally be involved.

Where the difficulties of perceiving contingency arise solely due to calculational requirements, the difficulties are explicit. But difficulties need not be so apparent. One might maintain that a statement had a complex logical structure, the statement in some way being (at least roughly) equivalent to one whose complex logical structure is explicit. The difficulties of perceiving the statement's contingency (or necessity) would then be double. First the basic structure must be brought to light, and then the calculations must be performed. An easier alternative method would be to provide a counter-example to the statement. This method might be thought of as (in some strange but intuitively understood way) very like falsifying the necessity of an explicit and logically complex statement, by the use of a tricky technique that avoided most of the calculations that would normally be required. In any case, it appears that the position that all necessary truths are analytic statements would lead us to something of a logical atomism in order to explain the *philosophical* plausibility of the refuted Empiricist doctrines. Along with a relatively atomistic theory of reference and cognitive meaning, one would import an accompanying psychology of surprise to help with the explanation.

An alternative, and I believe better, explanation of the nature of the doctrines and of my arguments against them might be provided by a more liberal theory of necessary truth like that being developed by Professor Hilary Putnam. Putnam does not identify the necessary truths with the analytic

statements. Rather, he thinks that the number of the former is greater.⁵ He puts forward reasons for this thought in his general views on the subject. I dare not expound Putnam's views on this subject except very sketchily and with apologies: on Putnam's view, a sentence expresses a necessary truth if, granted that it is unambiguous, the sentence expresses a statement which is understood in such a way that no empirical model can be recognized (by the understander) as a possible disconfirmer of the statement. Any empirical situation, as regards the domain talked about by the statement, can be interpreted in terms of the statement. And no empirical model will be understood in terms of a statement that contradicts the original one. The sentence is then accepted with the status of expressing a necessary truth. It has not been accepted because "its" statement has been confirmed by those empirical models that actually do obtain, the statement thus triumphing over conflicting hypotheses. Rather, the sentence is accepted without alternative.

I then suggest that it is initially reasonable to consider the Empiricist doctrines as necessary in a sense like Putnam's. It is initially reasonable to believe that the formulas expressing the doctrines might be accepted without alternative. They would express statements which no empirical model would be accepted as disconfirming. It might be, as the sceptic suggests, that if one really understood the import of the doctrines, one would understand that any empirical model would conform to their dictates; the statements expressed by the formulae are not reasonably to be understood as being disconfirmable by empirical models. But by way of counter-example, I have shown that this initially reasonable belief is not a correct one. Ultimately, then, it is not reasonable to believe that these Empiricist doctrines are necessarily true.

The exhibited examples do not allow us to accept the sceptical replies (A) and (B). For they speak most clearly and directly to the question of necessary truth, leaving no room for either sceptical maneuver or doubtful decision. Thus their impact is direct. And, of course, this impact can be explained: When one is discussing the Empiricist doctrines, one becomes wholly involved with the "purely psychological" terms used to express the doctrines. It is difficult in this context

to remind oneself that available for application to the questions at hand is the conceptual apparatus, e.g., statements I and II, which expresses a world view familiar to us.

In terms of this world view, descriptions of persons which employ only psychological terms are true in virtue of the truth of unspecified descriptions of persons which do not employ these terms but which do employ physical or physiological terms. (The perhaps necessary inability of providing an adequate translation of descriptions of the first kind in terms of descriptions of the second kind is completely irrelevant to the questions we are discussing.) In this paper I am drawing our attention to this particular conceptual view, partially expressed by statements I and II, and showing how the view allows for the construction of a situation which provides a clear counter-example to the Empiricist doctrines. In this way I show how given the conceptual apparatus we do have, we can construct a case which would not call forth the replies (A) and (B) and which, granting the truth of this view, renders these replies futile.

In terms of this world view, I show myself in the most definite way both what it would be like for a person to have fully developed (empirical) concepts although lacking previous experience, etc., and how it might come about that there be such a person. By reminding myself of statements like I and II in the midst of a discussion of the Empiricist doctrines (which are expressed in "purely psychological" terms), I can show myself that these doctrines are necessarily true only in case the statements expressing a familiar world view, e.g., I and II, are necessarily false. But *clearly* these statements, e.g., I and II, are *not* necessarily false, no matter what one's theory of necessity (so long as the theory is a reasonable one). Most generally speaking, in the (philosophers') sense of necessity, where necessity is strictly involved with truth and falsity simply in virtue of conceptual understanding (but not necessarily simply in virtue of the meanings of expressions or the rules of language), there is no reasonable theory of necessity on which the Empiricist doctrines are necessarily true. Indeed, this is what we show in our main refuting argument.

Now that we better understand the nature of

⁵ For Putnam's positive views on necessary truth, see his excellent essay "The Analytic and the Synthetic," in vol. III of the *Minnesota Studies in the Philosophy of Science* ed. by H. Feigl and G. Maxwell (Minneapolis, 1962), pp. 358-397. For reason to think that the occurrence of analytic statements (or, perhaps, more properly, analytic sentences) is not nearly so great as is often supposed, see Putnam's "Dreaming and Depth Grammar" in *Analytical Philosophy*, ed. by R. J. Butler (New York, 1962).

our problem and the nature of the main refuting argument which solves it, we go on to present several briefer sections where we pursue this understanding in further directions.

9. We can appreciate the simplicity and power of the argument of Section 5, which I shall now call the main refuting argument, by noting that it affords us with a workable counter-example which is wholly intelligible to us in terms of our present conceptual understanding. We do not need a more detailed understanding of the concept of a concept or of the concept of experience. We can rely solely on the fact that we do have an understanding of what it is to have a concept and of what it is to have experience. A more detailed understanding of these concepts would better enable us to answer questions involving judgments of similarity. But lacking detailed understanding, questions such as "Is x similar enough to y so that it is true that they both have (essentially) the same concept of C ?" usually receive answers that are somewhat arbitrary. Usually the case presented requires one to make a judgment of similarity that one is not well equipped to make. Not so in the case of our main refuting argument. The technique of the argument is to employ a case of exact physical, mental, and behavioral duplication. So questions of judgments of sufficient degrees of similarity are ruled out by the technique. One might say that not only is the case presented the paradigm one; but also that it is necessary that it be so.

10. Simplification is achieved in other directions as well. I need not consider whether congenitally blind men can have a concept of red (or of color, etc.), and so become involved in questions such as "What is meant by 'congenitally blind' and what does the use of this expression entail?" Nor need I become involved in the difficult questions of how to tamper with the brain of a mature person so that he would come to have color concepts, etc. And consequently, I need not become involved in the difficult questions of personal identity which this business of tampering might involve. For there are other cases than that of the congenitally blind man to which we can turn when looking for someone who has not had experience of a certain kind. As we have seen, one way of doing this is by turning to the case of a person who has had no experience at all.

11. Finally, my procedure of "blindly" copying a person allows me to construct a person with concepts without our having to figure out a recipe for how to proceed to accomplish this very generally

phrased requirement. Certainly there remain problems of information (how to find out about a living brain) as well as problems of construction. But these are not problems which, in this context, are of philosophical interest. Here, too, the power of the technique lies in its simplicity.

12. The example involved a situation where the duplicate person was constructed, but this is not essential. We can conjecture that the various parts of the physical assemblage of the duplicate person just ("by chance") came together. This is possible even in a world correctly and adequately described, say, by Newtonian physics. But, I need not even have a case where the coming-into-being of the duplicate person is predictable. It may be unpredictable, as well as unplanned, and the argument will still work. For the question simply is whether one would say of such a duplicate person that he did have fully developed concepts, but did not have (previous) experience. And the answer is, patently, that one would.

One might say that the cases of this section depend on an imagined sequence of events that is so unlikely as to be fairly called "impossible." To this I must say that "impossible" as used here virtually means the same as "unlikely in the extreme" and, so, this objection has nothing to do with the intent or meaning of the Empiricist doctrines under discussion, where a stricter necessity is claimed.

13. A natural reaction to my method of proof is to object: "But you just built the experience into the person." Now, I can't say that I know what building experience into someone is supposed to be. I think that I do know what building a capacity into someone is; indeed, I think that in the case of my proof I have built the capacity to employ certain concepts into someone. (I did not take someone and build this capacity into him; rather I built it in in building him.) I would have thought that experience is not the kind of thing which can be built into someone. Saying that we can build experience into someone may be an enlightening remark, but it should be recognized as a recommendation to change our concept of experience, or as a metaphor. This is because "experience" is used to talk about episodes; "concepts" is used to talk about capacities. With a sufficiently altered conception of experience we may, upon detailed examination, find that we can retain the Empiricist doctrines which we have been discussing. In truth, since we are using our actual concept of experience, we can say that all

we have done is to build someone as if he had certain experience; i.e., just as he would be if he had certain experience. But he did not have this experience. And it is not essential to the case, *once we understand it*, to consider that there was someone who did have experience and who served as a model for the construction. The second man would have the concept of red even if the first man had never existed.

14. As the phrase "had a previous experience" involves reference to a previous episode, so the phrase "learned something" involves an implicit reference to certain kinds of things having gone on and not others as concerns the person of whom it is said that he has learned. The *constitution* of a person's neural structure by the direct physical manipulation of a biologist is not a case of someone having learned something. I see no reason to think that this should or can be called a case where someone has *learned* something, this word having anything like its present meaning. For learning requires a learner, and a learner is a *more or less* complete, or responsive being. Given our present concept of learning, I do not see how one could successfully claim that learning had transpired in the case of our example.

15. The concepts of having concepts and of understanding a language refer to capacities. When a person has a certain concept or understands a language, he knows how to do certain things; he is able to do them. (Though perhaps the things are entirely mental in character.) I wish to maintain that there is no retroactive entailment involved in the application of either of these concepts, which entailment would effect the force of any of the examples that I set forth. Some phrases, for example, "is educated," "is practiced," "is experienced," etc., do involve implicit reference to previous occasions on which the person in question was being educated, was becoming practiced, was gaining experience, etc.⁶ I wish to *contrast* the phrases "has a concept of *x*" and "understands a language" with such phrases.

"Understands a language" and "has a concept of *x*" denote present capacities. I can see no reason to think that these phrases involve references to other facts involving the past, the present, or the future *which could affect the force of our argument*.

16. When we consider the case of someone just coming into being (however impossible this may

be from a scientific point of view) we see that the doctrine that concepts must be acquired is not true as a philosophical doctrine unless so interpreted so as to be entirely vacuous. In the case of someone being constructed one may say that he acquired the concepts in the process of his construction. In saying this one is saying something. One is saying that the person, as he is, did not just come into being ("particles and all"). And one is saying that it is not just an accident that he has the concepts. But if someone suddenly did just come into being ("particles and all") with a set of fully developed concepts, then it would not be the case that he acquired these concepts in some process going on. For there is no process. One may say that he acquired the concepts in his coming into being. But then I am not sure that one is saying anything at all. And I know that I, for one, would prefer not to make such a statement.

17. Once, by way of these examples, we have broken our initial general prejudice against the possibility of building things that have concepts just as we do, we can let other cases, which are not exact duplicates, count as cases where constructed things (persons here count as things) have concepts and do not merely "have concepts."

18. As a point of additional interest we may note an application of our proof in epistemological argument.⁷ It is most interesting to note that there is a conflict between the view that holds that for all we know there might not be a past world; i.e., everything in the world including ourselves and our "memories" came into being a moment ago and proceeded thenceforth in an orderly manner, and the Empiricist doctrines we have been discussing. For if our understanding depends on our having *prior* experiences or learning something, then the premiss that there is no past world, or more especially that we have no past life, must lead to the conclusion that we cannot understand this very premiss. The argument for scepticism concerning memory cannot be understood as an argument if the hypothesis that we have had no past is correct. The argument does not exist for us; all that exists for us is a series of inscriptions, utterances, or sensations. If what is maintained as a viable alternative, namely that we have had no past, were to be the case, then we could not understand the argument for this being a viable alternative. Given the truth of the

⁶ These phrases are suggested by Professor Gilbert Ryle.

⁷ This interesting connection is suggested by Saul A. Kripke.

Empiricist doctrines,⁸ scepticism concerning memory is not refuted in the sense that it is proved impossible that we have no past; but its status as a rational philosophical position may be considered as refuted. But the Empiricist doctrines are not true; they are false. Hence, *prima facie*, it is possible that we just came into the world, indeed the world (save if you will a single clock) just came into existence; and that we be able to understand arguments to the effect that this may be so. Why this must be so should be obvious to anyone who has understood and accepts the previous parts of this essay. Citing these Empiricist

doctrines as arguments against scepticism from memory is no good, for the doctrines must be rejected before the scepticism.

19. Finally, I think that it is obviously true that in the case of our own lives, experience has had an instrumental role in the development of our understanding. For us to get an adequate idea of how our experience has influenced the development of our understanding, it is certain that we would need a far more detailed knowledge of the concept of experience. The essentially private, non-spatial entity of dualism would seem not to get us very far toward this goal.

University of Wisconsin

⁸ And given the premiss that a person's present moment experience could not account for his having a concept of the past.

VI. CAN COMMANDS HAVE LOGICAL CONSEQUENCES?

G. B. KEENE

IN a recent paper,¹ Bernard Williams makes out a case for saying that there cannot be a logic of commands, in the sense of inference from commands, while at the same time claiming that there can be logical relations between commands. In reply to this,² P. T. Geach urges, in effect, that if there can be logical relations between commands, then there are *ipso facto* inferences from commands to commands. I shall argue (in agreement with Williams) that there cannot be any such thing as inference from commands to commands, nor (in disagreement with him) logical relations between commands either. In fact, my conclusion will be that: (a) there can be a relation between commands, but that it is not the relation of logical entailment, and (b) that there can be a "logic of commands," but that it is not one in which inferences can be made from commands to commands.

In the first place, to ask whether there can be logical relations between *statements* immediately strikes one as odd. For it would be like asking whether there can be mathematical relations between numbers. What else could such relations hold between? Now, in a sense of course, there can be mathematical relations between entities other than numbers. For example, there can be mathematical relations between the planets. To say this, however, is to say no more than that the movement, size, or some other measurable features of the planets stand in certain mathematical relations. Clearly we cannot insert "Jupiter" for the variable in any mathematical relationship. In short, mathematical relations are not, literally, *between the planets*; they are between mathematically expressed "descriptions" of the planets. (For example: " x [= the speed of Jupiter] < y [= the speed of Mars]."

Analogously, it is in order to claim that there can be logical relations between entities other than statements, for example the moves of a game such

as chess (or between different switching circuits). But here again we have simply a shorthand way of referring to relations between statements about pieces in the game. For example, " $Q-Q5$ " entails "ch" means, in full, something like:

$$B_1, \dots, B_k, Q-Q5 \quad \vdash_{A_1, \dots, A_n} \text{ch}$$

where B_1, \dots, B_k are assertions describing the positions of the pieces on the board at the time, A_1, \dots, A_n are the rules of play, and \vdash_{A_1, \dots, A_n} is the logical relationship of derivability relativized to a logic having A_1, \dots, A_n as additional rules (suitably re-expressed in the formal mode).

There is, however, a closer analogy that can be drawn between the domain of mathematical relations and the domain of logical relations. For just as talk about mathematical relations between planets derives from formulae concerned with mathematical relations (in the literal sense) between numbers, so talk about logical relations between statements derives from formulae concerned with logical relations (in the literal sense) between sentences which can be used to make true or false statements. To explain what I mean by this, I shall use the following abbreviations:

- (1) $P(x_1, \dots, x_n)$ FOR the indicative sentence P containing (usually implicitly) the context-relative variables x_1, \dots, x_n .

E.g., a particular occurrence of the sentence "The house is on fire" in which some specific house at some specific time is being referred to, would be understood to contain the appropriate context-relative variables, thus: "The- (x_1) -house is- (x_2) -on fire."

In other words, indicative sentences can be regarded as containing gaps which are most usually filled implicitly by the context of utterance (in other words not filled at all except that both

¹ "Imperative Inference," *ANALYSIS Supplement*, vol. 23 (1963), pp. 30-36.

² *Ibid.*, pp. 38-39.

speaker and hearer could and would fill them alike). Such inscriptionally unrecognized features of indicative sentences are what I am referring to as "context-relative variables."

- (2) $P(x_1, \dots, x_n)a_1, \dots, a_n$
FOR the sentence $P(x_1, \dots, x_n)$ with
 a_1, \dots, a_n as (explicit) values of the
context-relative variables
- (3) $[P(x_1, \dots, x_n)a_1, \dots, a_n]$
FOR the set of all utterances of
 $P(x_1, \dots, x_n)a_1, \dots, a_n$
- (4) P' FOR the indicative sentence P with all
its context-relative variables made
explicit
- (5) S/P FOR an utterance of the indicative
sentence P'
- (6) ct FOR "commits the speaker to accepting
the truth of"

Using this terminology, the way in which logical relations between statements are connected with logical relations (in the literal sense) between sentences which can be used to make true or false statements can be expressed as follows:

If S/P and S/Q are any members of the sets constructed from two indicative sentences P and Q respectively, in accordance with (3) above, then:

If $P \vdash Q$ then $S/P \text{ ct } S/Q$

Or, in full: If P logically entails Q , then an utterance of a context-definite version of P commits the speaker to accepting the truth of an utterance of the corresponding context-definite version of Q (where \vdash is relativized to a given standard first order function calculus).

Four points to note in connection with this implication are:

(i) The implication is one-way only. That is to say we could not claim:

If $S/P \text{ ct } S/Q$ then $P \vdash Q$

since the relation of *being good inductive evidence for* is also sufficient for $S/P \text{ ct } S/Q$ to hold.

(ii) We do not have:

If $P \supset Q$ then $S/P \text{ ct } S/Q$

For the notion of commitment here is that of logical commitment, where (a) a person X can be committed (by what he says) in virtue of a logical truth, whether he has uttered it or not, but (b) X cannot be committed (by what he says) in virtue of a factual truth which he has not uttered.

(iii) Being committed to accepting the truth of a proposition P (i.e., being committed to accepting the propriety of uttering P) is quite distinct from being committed to uttering P . For example, being paid as a clergyman would commit me to uttering, on certain occasions, "I believe in the Father, the Son and the Holy Ghost," but does not commit me to accepting the truth of this proposition. Otherwise, "I believe because I am paid to" would not be a joke.

(iv) It may be asked: How does a formalization of "commitment to accepting the truth of" differ from a formalization of orthodox first-order functional logic? For, it may be argued, surely every theorem in the one can be transcribed by a mere notational change into a theorem of the other, and *vice versa*?

The reply to this is implicit in what I have already said. Logical entailment between P and Q is not a necessary condition of $S/P \text{ ct } S/Q$. Thus for a well-defined concept of "good inductive evidence for" (say "ev"), we have:

If $P \text{ ev } Q$ then $S/P \text{ ct } S/Q$

Furthermore, for P = "I accept the truth of every statement made by X " and Q = "Killing innocent people today is justified if it results in more people, yet to be born, being saved from poverty," where Q is a statement made by X , it would seem reasonable to assert:

$S/P \text{ ct } S/Q$

even though it is neither the case that $P \vdash Q$, nor $P \text{ ev } Q$.

I turn now to a closer look at the concept of "logical relation between commands." In the first place it seems reasonable to claim that the *giving of* command A might be said (in some sense) to entail the *giving of* command B , or that the giving of command A was "command-inconsistent" with the giving of command B , in some sense of "command-inconsistent" yet to be specified. It should not therefore be altogether out of the question to set up a logic of commands in a manner parallel to that in which we have treated the logic of statements. The steps would be as follows:

- (i') $A(x_1, \dots, x_n)$ FOR the imperative sentence
 A containing (usually
implicitly) the context-
relative variables $x_1 \dots$
 x_n .

E.g. A particular occurrence of the sentence "Take it back!" in which the addressee, the object in question, and two places are implicitly referred to, would become context-definite in the form: " (x_1) -take it- (x_2) -back from- (x_3) -to- (x_4) "

- (2') $A(x_1, \dots, x_n)a_1, \dots, a_n$
FOR the sentence in question made context-definite by a_1, \dots, a_n
- (3') $[A(x_1, \dots, x_n)a_1, \dots, a_n]$
FOR the set of all utterances of the context-definite sentence given
- (4') A' FOR any context-definite imperative sentence
- (5') C/A FOR an utterance of the context-definite imperative sentence A' .

To complete the analysis, we have to find a counterpart to "commits the speaker to accepting the truth of" (in connection with statements). It could be argued that the required parallel is "commits the hearer to obey." For it is certainly true that, just as statements are not normally made when their truth is not accepted by the speaker, so commands are not normally given when the authority to give them is not accepted by the hearer. This suggestion, however, breaks down immediately. For, even given a relation of "entailment" holding between imperative sentences (as a counterpart to logical entailment) we could hardly claim:

If C/A and C/B are any members of the sets constructed from A and B respectively, in accordance with (3') above, then:

If A "entails" B then C/A commits the hearer to obey C/B . Or, in full: If imperative sentence A entails (in some supposed well-defined sense) imperative sentence B , then an utterance of a context-definite version of A commits the hearer to obeying an utterance of the corresponding context-definite version of B (i.e., the utterance of an imperative sentence having "corresponding" values for its context-relative variables, by the same speaker to the same hearer). This, as it stands, is absurd since the giving of a command does not, in itself, commit anyone to obeying that command, let alone to obeying some other command. On the other hand this implication comes near to formulating something not so absurd, namely—that if executing command A presupposes some action say a (not yet carried out) which has

not been commanded, then the giving of command A commits the speaker to the propriety of issuing a further command (to do a). I shall return to this point.

So far the position we have arrived at is this: even assuming that there can be a logical relation between command-sentences, the relationship between actually given commands (corresponding, in the case of statements, to the relationship between actually asserted sentences, defined on the relation of logical entailment between sentences) cannot be the relation: "commits the hearer to obey." Giving a command might in some circumstances be said to commit the speaker to seeing that it was obeyed but cannot significantly be said to commit the *hearer* to anything. Nor is it possible to argue that giving a command could in some circumstances be said to commit the speaker to giving some other command. For example, although the command "Obey the command that I shall give you in ten seconds from now" commits me to giving a further command, it does so only in the sense in which, if I say "I am going to hit you in ten seconds" I am committed to hitting you. But this is not the sense of "commit" which is required in the present context. For we require a sense of the word which is defined on the alleged relation of entailment between the imperative sentences in question. In short, we should want a situation in which C/A commits the speaker to C/B because A (in some sense) entails B . But what sentence does "Obey the command that I shall give you in ten seconds from now" entail? We cannot say that it entails the sentence which will be used in the command which (ten seconds ago) I said I would give you, for we do not (at the time) know what this sentence is. In any case the relation of entailment between imperative sentences is obscure and undefined. (I shall later find fault with one attempt to eliminate the obscurity in this notion.)

To improve on the analysis we must look for a better counterpart to "commits the speaker to accepting the truth of," for actually given commands. What is required here is some concept which plays a role in connection with commands, parallel to that played by the concept of truth in connection with statements. Suppose we had found such a concept C , we could then express, as our relation between actually given commands: "commits the speaker to accepting the C of." Now the truth of a *statement* amounts, in effect, to the propriety (with respect to the conventions of

statement-making) of making that statement.³ To make a false statement, whether wittingly or not, is to do something which is in contravention to an accepted code, or implicit conventions. It is part of an accepted code of linguistic behavior that the use of "stating language" be confined to the conveyance of information as to actual states of affairs. Only because this is so, can "stating language" be successfully misused for the purposes of deliberate deceit. Furthermore, there could be no such accepted code governing "stating language" were there not implicitly understood conventions of word-use enabling the hearer to read-off from the statement some expectancy with regard to what he would observe under certain conditions. Only because this is so can anyone be mistaken in what he states. If there were no conventions determining, for the hearer, one particular expectancy rather than another (from a given sound sequence) with regard to what he would observe under certain conditions, the hearer could not *correct* the speaker. A kind of speech performance which cannot be corrected cannot be used to deceive. So that if, by "improper" we mean "in contravention to an accepted code or of implicit conventions (or both)" then we can equate the truth of a statement with the propriety (with respect to the conventions of statement-making) of making that statement. In the case of commands there is the exactly parallel concept: the propriety (with respect to the conventions of command-giving) of giving a certain command. Just as a false statement is such that when once one is aware of the contravention involved one could not be expected to accept it, so an improper command is one that has features such that one who is aware of them could not be expected to obey it. Conversely, just as a *true statement* is one which it is in order (with respect to the conventions of statement-making) to make, so a *proper command* is one which it is in order (with respect to the conventions of command-giving) to give. We have, therefore, as our relation between actually given commands: "commits the speaker to accepting the (command)-propriety of." Using *CP* as our abbreviation for this relation, the implication corresponding to our previous one for statements would be:

If $A \vdash B$ then $C/A \text{ CP } C/B$

Or, in full: If the imperative sentence A logically

entails the imperative sentence B , then giving the command A commits the speaker to the propriety of giving the command B . However, since the relation of logical entailment does not hold between imperative sentences, we cannot accept this as our counterpart implication.

It might be argued that the implication could be modified to cope with this difficulty, by introducing

@ A FOR a description of the state of affairs that would result were C/A to be obeyed (the actualization of A')
and reformulating the implication as:

If $@A \vdash @B$ then $C/A \text{ CP } C/B$

That is: If a description of the actualization of the command A logically entails a description of the actualization of the command B then giving the command A commits the speaker to the propriety of giving the command B . But the trouble with this is that there are many entailments in respect to which we should not want to claim that the relation *CP* held between the corresponding commands. For example, since "He handed in his pass" entails "Either he handed in his pass or he shot the Captain" we would, by the above implication, have:

Giving the command "Hand in your pass" commits the speaker to the propriety of giving the command "Hand in your pass or shoot the Captain."

This is an unwanted result, if only because it would allow that being authorized to carry out some particular action would, by this logic, entail being authorized to do anything. For to say that I am committed to accepting the propriety of *giving* the command B , in virtue of having given the command A , amounts to admitting that I am committed to accepting the propriety of command B' being *executed*, if it were given. In short, I am committed to accepting the propriety of an action being carried out which would constitute the obeying of command B had it been given. To be so committed is to have implicitly authorized this action by having explicitly authorized the carrying out of command A . But P entails $(P \vee Q)$ for any Q . Hence, authorizing the actualization of some command-sentence A is implicitly to authorize the actualization of either A or command-sentence B , for any B .

³ The illustration which follows clearly presupposes a detailed analysis of "truth-conditions" which is built-in to the notion of "propriety" in this context.

It is difficult, at first sight, to see what has gone wrong here. Clearly we would want to say that, in general, when obedience of person X to one command is not possible without person X carrying out some particular action a (where a is not mentioned in the command and has yet to be carried out), that command implicitly authorizes the carrying out of action a . For example, if it is not possible to report to the guard-room without passing the officers' mess, then the command "Report to the guard-room" carries with it the understood authorization to pass the officers' mess. In these circumstances "Report to the guard-room" stands in the relation CP to giving the command "Go past the officers' mess." Yet it will not be the case that "He reported to the guard-room" logically entails "He went past the officers' mess," since he could have flown there.

What this suggests is that actions can stand in a relation to one another independently of the logical relation holding between assertions that the actions in question have taken place. This relation is not explicitly recognized in ordinary speech, unless by use of that maximally vague word "involves." What is wanted here is a logic of *actions*, in which a well-defined concept of *inclusion* plays a leading role. Let us suppose we have such a logic. It might well contain as an axiom or theorem (using $!$ for "the action of obedience to" and $incl$ for "inclusion"):

If $!C/A \text{ incl } !C/B$ then $S/@A \text{ cr } S/@B$

but not:

If $S/@A \text{ cr } S/@B$ then $!C/A \text{ incl } !C/B$

For we already have:

If $P \vdash Q$ then $S/P \text{ cr } S/Q$

and we do not want:

If $S/@A \vdash S/@B$ then $!C/A \text{ incl } !C/B$

since, for example: "He handed in his pass" logically entails "He has handed in his pass or he has shot the Captain," but the action of handing in one's pass does not include the action of shooting the captain.

On this basis, our axiom of command propriety may be formulated as follows:

If $!C/A \text{ incl } !C/B$ then $C/A \text{ cr } C/B$

Or, in full: If obedience to an utterance of a context-definite version of the imperative sentence

A includes obedience to an utterance of the corresponding context-definite version of the imperative sentence B , then giving the command A commits the speaker to the propriety of giving the command B .

A realistic example of the application of the axiom of command propriety is to be seen in the Supreme Court ruling in the case of *Edward's Lessee vs. Darby*.⁴ The dispute concerned a North Carolina statute which expressly required the commissioners to have surveys made of certain lands granted to soldiers. The lands in question were located around some salt licks and springs which were not subject to the act establishing land grants for soldiers, and the reservations around which were to be avoided. The Court said:

We admit the statute does not give the authority to survey the reservations, in express terms, but do not admit that the authority may not, and does not, result by necessary implication from the duties they were expressly required to perform. . . .

Although the statute did not expressly require the commissioners to determine "What licks and springs were proper objects for reservation," nevertheless since it did require them to survey the lands granted to the soldiers, in such a way as to avoid these reservations,

It seems to result necessarily from these provisions, that the commissioners must first determine what were the proper subjects of reservation, and having determined that a given salt lick or spring came within the provisions of the law, the power and duty of laying off by survey the 640 acres reserved, and to be avoided, around the lick, seems necessarily and irresistibly to result to the commissioners, in all cases where they might deem it necessary to do so, in order to enable them to lay off the lands for the officers and soldiers, so as to avoid these reservations. . . .

Here the Supreme Court is arguing, in effect, that since the act of surveying certain lands containing salt licks and springs to the exclusion of the salt licks and springs includes the act of surveying the salt licks and springs themselves, an express command to survey the lands commits those issuing that command to the propriety of issuing a command to survey the salt licks and springs. Since being committed to accepting the propriety of issuing a command (without actually doing so) is tantamount to authorizing the actions

⁴ Quoted in Hans J. Morgenthau, *Dilemmas of Politics* (Chicago, 1958), pp. 146-147.

which constitute obedience to that command, the commissioners are, by implication, authorized to survey the licks and springs.

Thus the present analysis finds room for a relation between commands parallel to that which holds between the making of statements. But it shows at the same time that all talk of inference in connection with commands is parasitic upon inference in the straightforward logical sense. For the axiom of command propriety has the limited force of linking relations between commands to statements asserting relations of inclusions between actions. We can, for example, infer from the giving of a command, in certain circumstances, to the propriety of the speaker's giving some other command. In this sense, we have a logic of commands, but it is not a logic in which we can infer a command from a command.

I will now try to summarize the position which I have been trying to maintain. Let *LC* be the following position: Just as there is a relation between actually uttered indicative sentences which can be defined in terms of the relation of entailment between indicative sentences themselves, so also there is a relation between actually uttered imperative sentences which can be defined on the relation of entailment between imperative sentences themselves. *LC* clearly depends on the acceptability of the following position: There can be an entailment-relation between imperative sentences. (Call this *EC*.)

LC involves a distinction between inscriptions and utterances and correspondingly between relations between inscriptions and relations between utterances. (The difference between utterance and inscription, here, is not of course merely that the one can be heard and the other not. Utterance, here, means something more like "issuing to an intended audience.") Furthermore, *LC* requires there be a syntactic relation between those inscriptions that are imperative sentences, which coincides with the syntactic relation (of logical entailment) holding between those inscriptions which are indicative sentences.

Perhaps the most outspoken defendant of entailment between commands is R. M. Hare in *The Language of Morals* (Oxford, 1961). His argument there (p. 25) is this: A sentence *P* entails a sentence *Q* if and only if the fact that a person assents to *P* but dissents from *Q* is a sufficient criterion for saying that he has misunderstood one or other of the sentences. Suppose someone assented to "Take all the boxes to the

station" and also to the statement "This is one of the boxes" and yet refused to assent to "Take this one to the station." This would be enough for us to conclude that he had misunderstood one of these three sentences. Thus there are entailment-relations between commands.

The trouble here is with the word "assent." Assenting to "Take all the boxes to the station" means *agreeing to obey* the command; assenting to "This is one of the boxes" means *accepting the truth* of the statement. Clearly, the definition Hare gives of entailment uses the word "assent" ambiguously as between these two senses. When the ambiguity is removed, the definition we are offered runs.

An (imperative) sentence *P* entails an (imperative) sentence *Q* if and only if the fact that a person agrees to obey *P* but refuses to obey *Q* is a sufficient criterion for saying that he has misunderstood one or other of the sentences.

This says, in effect, that on any occasion on which a person's having agreed to obey some command *P*, but refused to obey a further command *Q*, is taken as showing that he must have misunderstood one of the commands. On any such occasion we can say "*P* entails *Q*." The crucial question here is "In what sense of 'entails'?" Certainly not in the sense:

P entails *Q* if and only if the fact that a person agrees to accept the truth of *P* but refuses to accept the truth of *Q* is a sufficient criterion for saying that he has misunderstood either *P* or else *Q*.

For this sense of "entails" is tied strictly to indicative sentences. In short, that we can say "*P* entails *Q*" in the circumstances in question is simply what the definition stipulates: call this sort of situation a case of entailment between commands.

Hare not only concludes, unjustifiably, that there can be entailments between commands, but also that: "Whether the word 'entail' is used for these relations is only a matter of terminological convenience" (p. 26). It is, on the contrary, a matter of great terminological inconvenience, since "logically entails" is tied to "accepting the truth of," but not to "agreeing to obey," in much the same way as "square root of" is tied to "the number . . ." but not to say "the word. . . ." It would be terminologically inconvenient to use the word "root" in a context in which reference was sometimes to be understood as the square root of a number and sometimes as the structural root of a word.

But there is a further objection to Hare's argument on this point. In his analysis of entailment in general, the condition: "assents to *P* but dissents from *Q* is a sufficient criterion for saying that he has misunderstood one or other of the sentences" does not cover exactly the details of his example of the command to take all the boxes to the station. For in applying his definition to this example, he says assent to "Take all the boxes to the station" and "This is one of the boxes" together with dissent from "Take this to the station," would be enough for us to conclude that he has misunderstood one of these *three* sentences. So that, with the ambiguity of "assent" removed, the definition we are offered is in fact:

An (imperative) sentence *P* entails an (imperative) sentence *Q* if and only if the fact that a person agrees to obey *P* but refuses to obey *Q*, while accepting the truth of some (indicative) sentence *R* (such that it would not be in order to give the command *Q* if *R* were not true) is a sufficient condition for saying that he has misunderstood one or other of the sentences *P*, *Q*, or *R*.

This definition is not only open to the same objection as we have just made to the original one, but it also brings out the point that assertions about relations between commands boil down in the end to assertions about the propriety of giving certain commands under certain conditions. In other words, the command "Take all the boxes to the station" entails "Take this one to the station"

in the sense that giving the command "Take all the boxes to the station" commits the speaker to the propriety of giving the command "Take this one to the station." Since "this one," being one of the boxes referred to in the original command, the action of taking all the boxes includes the action of taking this particular one.

I do not therefore reject *EC* as unsubstantiated but as incomprehensible. I therefore reject *LC*.⁵ On the other hand I accept the view that there is a pragmatic relation holding between actually uttered commands which is to some extent parallel to that relation ("commits the speaker to accept the truth of") which holds between actually asserted statements. This relation ("commits the speaker to accepting the command-propriety of") is, however, defined, not on *entailment* but on a so-far-unexplored relation of *inclusion* between actions. In consequence, statements asserting the incidence of this relation of inclusion between actions may well entail statements relating the actual giving of one command to the propriety of giving another command (or to the implicit authorization of some course of action not named in the original command).

My conclusion is: (a) there can be an "implication" relation between commands, but it is not that of logical entailment, and (b) there can be a logic of commands, but it is not one in which inferences can be made from commands to commands.⁶

University of Exeter

APPENDIX

A miniature proof that there cannot be a logic of commands could be given as follows:

Definition of Terms

"Command" for *imperative utterance to suitable audience*

"Entails" for *logically precludes the denial of Premises*

- (1) Wherever there is an inference from *A* to *B*, *A* can be said to entail *B*.
- (2) A command cannot be said to be denied.
- (3) Nothing can be said to preclude the denial of that which cannot be said to be denied.

⁵ See the Appendix.

⁶ For the elaboration, an approach to command-logic compatible with the present considerations, as well as a conspectus of the relevant literature, see Nicholas Rescher, *The Logic of Commands* (London, 1965).

Proof

- (i) Suppose there could be an inference from some command *A* to some command *B*
- (ii) Command *A* entails command *B* (by Premiss 1)
- (iii) Command *A* precludes the denial of command *B* (by definition)
- (iv) Command *B* cannot be said to be denied (by Premiss 2)
- (v) Command *A* cannot be said to preclude the denial of command *B* (premiss 3)
- (vi) Command *A* both can and cannot be said to preclude the denial of command *B* [by (iii) & (v)]
- (vii) There cannot be, in any sense, an inference from some command *A* to some command *B*.

VII. TEMPORAL PARTS AND SPATIO-TEMPORAL ANALOGIES

J. W. MEILAND

TO what extent is time similar to space? Richard Taylor is a leading exponent of the view that time and space are very similar; indeed, he claims that time and space are "radically alike."¹ I want to examine this claim by considering Taylor's argument for it, particularly with regard to his use of the notion of temporal parts.

The notion of temporal parts is also an essential feature of the so-called "manifold theory of time" as the following quotations show:

According to our analysis, a concrete phenomenal individual, ordinarily said to be in time, is regarded rather as having time in it. . . . What we think of as a phenomenal thing is distinguished from what we think of as a phenomenal event or process only in the pattern of differences among its *temporal parts*. A thing is a monotonous event; an event is an unstable thing. . . . Because time is linear, the shape of temporally continuous individuals varies directly with their size. . . .² (Italics added.)

Time "flows" only in the sense in which a line flows or a landscape "recedes into the west." That is, it is an ordered extension. And each of us proceeds through time only as a fence proceeds across a farm: that is, *parts* of our being, and the fence's, occupy successive instants and points, respectively. There is passage, but it is nothing extra. It is the mere happening of things, their strung-along-ness in the manifold.³ (Italics added.)

Accordingly, the conclusions which I shall try to establish concerning the significance of the notion of temporal parts will apply to the manifold theory of time as well.

¹ Richard Taylor, "Spatial and Temporal Analogies and the Concept of Identity" in *Problems of Space and Time*, ed. J. J. C. Smart (New York, Macmillan, 1962), p. 381. First published in *The Journal of Philosophy*, vol. 52 (1955). Page numbers in the text are references to the Smart volume. For a very different criticism from mine of the Taylor position and some interesting suggestions about how the term "object" should be used, see J. Jarvis Thomson, "Time, Space, and Objects," *Mind*, vol. 74 (1965), pp. 1-27. See also R. Taylor, "Moving About in Time," *Philosophical Quarterly*, vol. 9 (1959), pp. 289-301; W. J. Huggett, "Losing One's Way in Time," *Philosophical Quarterly*, vol. 10 (1960), pp. 264-267; B. Mayo, "Objects, Events, and Complementarity," *Philosophical Review*, vol. 70 (1961), pp. 340-361; F. Dretske, "Moving Backward in Time," *Philosophical Review*, vol. 71 (1962), pp. 94-98.

² Nelson Goodman, *The Structure of Appearance* (Cambridge, Harvard University Press, 1951), pp. 285-286.

³ Donald C. Williams, "The Myth of Passage" in *American Philosophers at Work*, ed. S. Hook (New York, Criterion Books, 1956), pp. 320-321. First published in *The Journal of Philosophy*, vol. 48 (1951), p. 463. See also W. V. Quine, *From a Logical Point of View* (Cambridge, Harvard University Press, 1953), pp. 65, 67.

I

The notion of temporal parts is essential to Taylor's attempt to show that there are many complete analogies between space and time. For example, one who denied that space and time were similar might contend that entities can move in space but cannot move in time. To show that there can be temporal movement which is very similar to spatial movement, Taylor first characterizes a case of spatial movement as:

At T_1 A is north of B
At T_2 A is south of B

and then characterizes the corresponding temporal motion as:

At L_1 A is future to B
At L_2 A is past to B

where T_1 and T_2 denote times and L_1 and L_2 denote places or spatial positions. But can any phenomenon satisfy this description of temporal movement? Taylor believes that this description can be satisfied:

Let A , for example, be an earthquake, occurring gradually over an area which includes two towns, and let B be a stroke of a clock (any place in the world). Now it is possible that in one town, A is future to B , and in the other, past to B . This fulfills the description. (P. 387.)

Again, one who believed that space and time are very different might contend that ". . . an object cannot be in two places at once, though it

can occupy two or more times at only one place" (p. 383). Taylor's reply is that an object can occupy two times at one place only if the object also occupies all the times between those two times. It can be at L_1 at both T_1 and T_4 only if it is at some spatial location at T_2 and T_3 also (where T_2 and T_3 are after T_1 and before T_4). (Presumably if the object does not occupy the times in between T_1 and T_4 , there will be a temporal "gap" in its existence; and the question will then arise as to whether the *same* object exists at T_4 as existed at T_1 .) If this is the correct description of the case of an object's occupying two or more times at only one place, then the analogous situation, according to Taylor, is an entity's existing in two places at one time while occupying all the places between those two places: "A ball, for instance, occupies two places at once, if the places be chosen as those of opposite sides; but in doing so, it also occupies all the places between" (p. 383).

But is the analogy complete? It might be said that the ball is not in two places at once. Instead, one *part* of the ball is in one place and another *part* in the other place, with still other parts in the locations between these two. Yet in the temporal component of the supposed analogy, the *whole* object is at T_1 and the *whole* object is at T_4 (and at the times between). Taylor's reply to this is:

It is tempting to say that only part of the ball is in either place; but then, it is a different *temporal* part of an object which, at the same place, is in either of two times. (P. 383.)

It is clear from this reply by Taylor that the use of the notion of *temporal part* is absolutely essential to at least the completeness of this putative analogy and perhaps to the existence of a significant analogy at all. This notion also plays a crucial role in most of the other putative analogies which Taylor describes. Consequently, it is essential that this notion of temporal parts be examined.

Taylor introduces the notion of temporal parts in the following passage:

Things can be spatially long or short, but so too they can have a long or brief duration, i.e., be temporally long or short. Indeed, there is no reason why temporal dimension should not be included in any description of the shape of a thing. The notion of length, in turn,

leads to that of *parts*, both spatial and temporal. Distinctions between the spatial parts of things are commonplace, but it is no less significant to reason that things have temporal parts too, often quite dissimilar to each other—for instance, widely separated parts of a man's history, or narrowly separated temporal parts of a kaleidoscope.⁴

This passage indicates that by "temporal part" Taylor means what we might call a "temporal cross-section." A temporal cross-section of an object may be represented in the following way: let two planes be passed through the line representing the temporal dimension, and let these planes be perpendicular to that line; these two planes intersect the line at different points, for example, at T_1 and T_4 ; then we can speak of a certain temporal part of that object, namely the part existing between those two planes. Another temporal part of the same object would consist of the cross-section between T_4 and T_7 . A third part would consist in the cross-section between T_3 and T_6 . Temporal parts can, of course, overlap one another, as the third of these parts overlaps the first two. The existence of many, if not all, putative analogies between space and time depends on the existence of spatial parts and temporal parts. For the components of these analogies make essential use of such parts. Therefore, the analogies between space and time will depend on and vary with the analogies between spatial parts and temporal parts. Insofar as there are disanalogies between spatial parts and temporal parts, there will be a lack of analogy between space and time. In the next section I discuss disanalogies between spatial and temporal parts.

II

What is a temporal part of a physical object? We have described it as a "temporal cross-section" of that object. Thus the object itself does not completely exist or is not completely present at any one very small interval of time. For the object is a *set* of temporal parts of itself and only one temporal part can be present during a given very small interval. Let us suppose that object X exists from T_1 to T_8 . Pass planes through the time line at T_3 and T_7 . This divides the object into three contiguous and non-overlapping temporal parts, namely (T_1-T_3) , (T_3-T_7) , and (T_7-T_8) .⁵ The

⁴ "Spatial and . . .," *op. cit.*, p. 382.

⁵ Temporal intervals will be represented by symbols of the form " T_i-T_j ." Temporal parts will be represented by symbols for temporal intervals in parentheses. Thus the temporal part of an object which exists during the temporal interval T_1-T_3 is represented by " (T_1-T_3) ."

physical object X may then be regarded as an ordered set of these parts, the order being that just given. Thus, if we use the notion of temporal parts, we cannot use the expression "object X " to refer to the object as it is in the interval between, for example, T_3 and T_7 . For what there is between T_3 and T_7 is only a *temporal part* of the object, not the object itself. One can no more use the expression "object X " to refer to what is in fact a temporal part of that object than one can use the expression "George's automobile" to refer to the engine of that automobile or any other spatial part of that automobile.

The expression "object X " in the preceding paragraph can be replaced by the name or definite description of any physical object; what was said there applies to all physical objects. But spatial parts of physical objects—for example, the engine of an automobile—are themselves physical objects. Hence everything that was said above about object X can be said about any spatial part of any physical object. Furthermore, spatial parts exist within time intervals, just as all physical objects do. Therefore, spatial parts of physical objects, being themselves physical objects, have temporal parts, just as all physical objects do. Spatial parts can be regarded as ordered sets of temporal parts too. So if object X perdures from T_1 to T_8 , and if X has a , b , and c as its spatial parts throughout that interval, then each of these spatial parts, as well as object X itself, will have temporal parts (T_1 – T_2), (T_3 – T_7), and (T_7 – T_8). For example, a will have a temporal part (T_1 – T_3), and b will have a temporal part (T_1 – T_3) different from but at least temporally similar to that temporal part of a .⁶

Now we must determine whether, and to what extent, there is an analogy of Taylor's sort between spatial parts and temporal parts. First of all, what sort of analogies does Taylor seek to establish between space and time? Taylor quotes Nelson Goodman as saying that "... a minimal spatially changing (moving) compound not merely occupies two places but occupies them at different times. ... Analogously, a minimal temporally changing compound would have not merely to occupy two times but to occupy them at different times." Taylor admits that this is absurd, but he does not regard this as indicating a serious disanalogy between space and time, for Goodman's formulation is regarded by Taylor as inadequate.

Taylor says: "But if this were a *real* analogy, the last word of this [Goodman's] statement would be 'places' rather than 'times,' in which case there would be no absurdity at all" (p. 395, n. 10; italics added).

This comment on Goodman's statement illustrates Taylor's procedure and shows what type of analogy he wants to establish between space and time. Taylor's procedure is to formulate the spatial component of the analogy and then to derive the temporal component of the analogy by replacing every spatial term in the spatial component by its corresponding temporal term and every temporal term in the spatial component by its corresponding spatial term. This procedure was illustrated in the first section of this paper. Taylor's comment on Goodman suggests that only such *complete* replacement of terms will give any analogy at all, let alone a complete analogy, for he indicates that only such a procedure will produce what he calls a "real" analogy.

Does the sort of analogy which Taylor requires exist between spatial parts and temporal parts? We have seen that *spatial* parts of a physical object have *temporal* parts. It seems that if there were an analogy of the *Taylorian* variety between the two sorts of parts, temporal parts of a physical object would have to have spatial parts. But it is certainly not the case that temporal parts of objects have spatial parts. The following example shows that temporal parts of a physical object do not have spatial parts. Let physical object X exist from T_1 to T_{10} . Let X be composed throughout this time period of two spatial parts called " a " and " b ." These spatial parts a and b are thus physical objects which have histories of their own between T_1 and T_{10} . Now, form the temporal part (T_3 – T_7) of object X by passing perpendicular planes through the time line at T_3 and T_7 . These planes enclose not only the temporal part (T_3 – T_7) of object X , but also the temporal part (T_3 – T_7) of spatial part a of object X and the temporal part (T_3 – T_7) of spatial part b of object X . A temporal part of X lies between these planes, but so also do these temporal parts of these two spatial parts of X . Surely only what lies between these planes can be counted as parts of the temporal part (T_3 – T_7) of object X . Yet all that lies between these planes, other than that temporal part of X , are temporal parts of spatial parts of X . The spatial parts of X do not themselves lie between these

⁶ Since a and b have different spatial locations, these temporal parts of a and b have different spatial locations or spatial properties and are therefore distinct.

planes. (These spatial parts lie between planes at T_1 and T_{10} as does object X itself.) Hence if temporal part (T_3-T_7) of object X has parts at all, its parts are *temporal* parts of spatial parts of X . What is present in this temporal part of the object is not the spatial parts themselves, but rather *temporal parts of those spatial parts*.

The expression "spatial part a " denotes an entity which exists in the interval T_1-T_{10} . The expression "spatial part a " does not denote that spatial part as it is during a sub-interval of its history. The entity which does exist during a sub-interval of the spatial part's history is not the spatial part itself but instead a temporal part of that spatial part. Hence the temporal part (T_3-T_7) of the whole object X consists of the temporal part (T_3-T_7) of X 's spatial part a plus temporal part (T_3-T_7) of X 's spatial part b . Thus, temporal parts or cross-sections of a physical object (in this case X) have *temporal* parts or cross-sections of spatial parts as their parts. Temporal parts of a physical object do not have spatial parts as their parts, as they would have to have if a Taylorian analogy were to exist between spatial parts and temporal parts. Hence no Taylorian analogy exists in this respect between temporal parts and spatial parts.

It is true that a temporal part of a physical object will have what might be called a "spatial aspect," "spatial properties," or even "spatial dimensions." For example, that temporal part will have a certain spatial size, it will be twenty yards from a temporal part of some other object, and so on, since, as we ordinarily say, "the object" has a certain size or is a certain distance from some other object in that temporal interval. And it might be claimed that this spatial aspect or set of spatial dimensions is part of the temporal part. So a temporal part might be said to have spatial "parts" of this sort.

But even if we do say that temporal parts have spatial "parts" of this sort, there is still a crucial dissimilarity between spatial parts of physical objects and temporal parts of physical objects. For spatial parts of physical objects are composed *wholly* of temporal parts, but temporal parts of physical objects are *not* composed *wholly* of spatial aspects. If a Taylorian analogy existed between spatial and temporal parts of physical objects, then both of the following statements would be true:

(S_1): A spatial part of a physical object can be regarded as a set of temporal parts.

(S_2): A temporal part of a physical object can be regarded as a set of spatial parts (of the sort known as "spatial aspects").

Statement S_1 is true. But what could statement S_2 mean? We might try to give S_2 a meaning in this way. Suppose that the temporal part of X in question is (T_3-T_7) . This temporal part has its own temporal parts, namely (T_3-T_4) , (T_4-T_6) , (T_6-T_7) and (T_3-T_7) . And each of the latter has a spatial aspect. So it might be thought that the temporal part (T_3-T_7) could be regarded as the set of these spatial aspects, that is, as the set of spatial aspects of its own temporal parts. This would provide a meaning for statement S_2 but it would also render S_2 false. For even with the set of these spatial aspects, there is still something missing from temporal part (T_3-T_7) , namely its *temporal* aspect. Hence the temporal part (T_3-T_7) cannot be regarded as composed *wholly* of spatial elements. This temporal part is not the set of certain spatial aspects.

I have been trying to show that there is at least one very important respect in which temporal parts are not analogous (in Taylor's sense of "analogous") to spatial parts. If these two types of parts are not analogous, then any putative analogy between space and time in which these two types of parts play an essential role is to that extent not an analogy. Most, if not all, of Taylor's alleged analogies between space and time have a spatial component and a temporal component, as the examples previously given show. And spatial parts play an essential role in one component while temporal parts play an essential role in the other component. So if, as I have tried to show, spatial parts and temporal parts are not analogous to one another, space and time will, to that extent, not be analogous to one another.

III

But there are much more crucial respects in which Taylor's analogies are in fact not analogies. In this section I want to show what these respects are, on the basis of the nature of temporal parts as described in the preceding section.

Let us first consider the case, already described in Section I, in which an object can be both past and future with respect to another entity. The spatial component of this alleged analogy is stated by Taylor as follows:

At T_1 A is north of B
At T_2 A is south of B

But this is an incorrect statement of the spatial component. For the expression 'A' is being used here to refer not to A itself, but instead to certain temporal parts of A, namely the temporal parts (T_1) and (T_2).⁷ Instead of saying that A is north of B at one time and south of B at another time, Taylor should instead say that a certain temporal part of A, namely N, is north of a temporal part of B and another temporal part of A, namely S, is south of another temporal part of B. So the correct statement of the spatial component of this alleged analogy is:

At T_1 temporal part N of A is north of temporal part G of B

At T_2 temporal part S of A is south of temporal part H of B

Given this reformulation of the spatial component of this alleged analogy, how is the temporal component to be reformulated? Taylor's original formulation is:

At L_1 A is future to B

At L_2 A is past to B

His example is that of an earthquake which occurs before a stroke of a clock in one town and after that stroke of the clock in another town. It is clear from this description that the earthquake is "spread out" in time. Therefore Taylor must say that a temporal part F of the earthquake occurred before the stroke of the clock and another temporal part P of the earthquake occurred after the stroke of the clock. So the correct formulation of this component of the analogy is:

At L_1 temporal part F of A is future to B

At L_2 temporal part P of A is past to B

(In this case, as in the case of the other component, the temporal parts of A have different spatial aspects or different spatial locations.)

It is clear from these reformulations that no "real" analogy exists here, in Taylor's sense of the expression "real analogy." For it is not the case that every temporal term in the spatial component has been replaced by a spatial term in the temporal component. The term "temporal" in the expressions "temporal part N" and "temporal part S" has not been replaced by the term "spatial." If the term "temporal" had been so replaced, we would have "spatial part F" and "spatial part P"

in the second or temporal component of the alleged analogy. And it cannot be replied that we can and should substitute "spatial" for "temporal" in this way. For what occurs at L_1 is not a spatial part of the earthquake. It is a temporal part of the earthquake because it occurs during a certain interval, with other temporal parts of the earthquake occupying other intervals. F and P are temporal parts of the earthquake though they have spatial aspects.

To take another example, Taylor claims that an entity can be at two times in one place, just as it can be in two places at one time (by being at all of the places between the two). When reformulated, this alleged analogy is:

At T_1 a temporal part of A is at L_1 and L_2

At L_1 one temporal part of A is at T_1 and another temporal part of A is at T_2

This is obviously not a Taylorian analogy, for the proper replacements of terms have not been and cannot be made.

Finally, Taylor asserts that things can move closer together or farther apart in time, just as they can do so in space (pp. 390-392). He characterizes this motion in space as:

At T_1 temporal-part₁ of object A and temporal-part₁ of object B are separated by a spatial interval X

At T_2 temporal-part₂ of object A and temporal-part₂ of object B are separated by a spatial interval Y, larger or smaller than X

The alleged analogue to this is:

At L_1 spatial-part₁ of object A and spatial-part₁ of object B are separated by a temporal interval X

At L_2 spatial-part₂ of object A and spatial-part₂ of object B are separated by a temporal interval Y, larger or smaller than X

(Notice that there Taylor does replace the "temporal" which occurs in "temporal part" in the first component with "spatial.") As an illustration of this second component, that is, of moving closer together or farther apart in time, he gives "... two rolls of thunder, considered as aerial disturbances either heard or unheard, each existing in two nearby towns, separated in one town by an interval in one second, and in the other by an interval of two seconds" (p. 391). What he seems to mean

⁷ The expressions " T_1 " and " T_2 " will be regarded as denoting very small intervals of time.

here is this: in the first town one roll of thunder is (or could be) heard one second after the other, while in the second town, it is heard two seconds after the other. Thus at T_1 the first roll is heard at L_1 ; at T_2 (one second after T_1) the second roll is heard at L_1 . And since these rolls are heard at different times at L_1 , what is heard is not a spatial part of each roll but instead a *temporal* part of each roll. Similarly, at T_1 the first roll is heard at L_2 and at T_3 the second roll is heard at L_2 . But again what is heard at L_2 in each case is a temporal part of each roll. So this second component of the analogy must be reformulated in terms of temporal parts, for example:

At L_1 temporal-part₁ of A and temporal part₁ of B are separated by a temporal interval X

At L_2 temporal-part₁ of A and temporal-part₂ (for T_3) of B are separated by a temporal interval Y larger or smaller than X

Again, it is clear that the alleged analogy does not meet the specifications set up for analogies by Taylor and hence does not count as a "real" analogy. For the term "temporal" as it occurs in the expression "temporal-part₁" in the first component is not now replaced by the term "spatial" in the second component. *Temporal* parts are involved in *both* components.

It might be replied to this that what has been shown by the above is that there are not complete or thorough-going analogies between space and time. The completeness of an analogy is to be measured by the extent to which spatial or temporal terms have been replaced by their corresponding temporal or spatial terms. But in the reformulated analogies, though replacement of terms is not complete, it is almost complete. And so the analogies between space and time might be thought to be very good analogies, though not perfect analogies. That is, Taylor's alleged analogies, when reformulated, show that time is very much like space though not exactly like space.

But if Taylor did reply in this manner, he would have to allow the disanalogy alleged by Goodman in the passage quoted in Section II. That disanalogy is one in which only one term in the first component is not replaced by its corresponding term in the second component, with the result that a serious disanalogy between space and time is exhibited. But if *analogies* having less than complete replacement of terms are proper and well-formed, then *disanalogies* (such as the one cited by Goodman) which have less than complete replacement

of terms are also proper and well-formed. So if Taylor allows less than thorough-going replacement of spatial and temporal terms, he will have to allow the existence of the disanalogy cited by Goodman and perhaps many other serious disanalogies between space and time. And it is probable that it would then not appear that time is similar to space to any great extent. This is perhaps why Taylor regards and must regard Goodman as not having followed the proper procedure in formulating analogies and thus as not having exhibited a "real" disanalogy. Taylor cannot allow less than thorough-going replacement of spatial and temporal terms and still claim that time is very much like space.

IV

This consideration of the notion of temporal parts not only shows that Taylor has not established the sort of analogies between space and time that he claims to have established, but also it shows what anyone who employs the notion of temporal parts must believe the relation between time and space to be.

In Section II we saw that temporal parts have spatial aspects. Temporal parts can have spatial relations to one another. For example, temporal part F of object A and temporal part F of object B are twenty yards apart if, during the time interval in question, what would usually be called "the object A " and "the object B " are twenty yards apart. The temporal parts have spatial properties or aspects. Hence the spatial aspect of a temporal part should be considered to be a *dimension* (or set of dimensions) of that part. We also saw that while temporal parts have other temporal parts as their parts, spatial parts have temporal parts as their parts. In fact, a spatial part can be regarded as a set of temporal parts. Both of these facts—that spatial *aspects* are dimensions of temporal parts and that spatial *parts* are sets of temporal parts—indicate a kind of priority of time over space. Spatial *aspects* are in some sense "dependent" on temporal parts, being dimensions or properties of temporal parts; and spatial *parts* are also in some sense dependent on temporal parts, being sets of temporal parts.

This priority of time over space is also shown by what was said in Section III. There we again found that it is temporal parts of objects which have spatial relations to one another. For example, an object which is at L_1 at T_1 and at L_2 at T_2 has

temporal parts which are at different *places* and hence which are spatially separated from one another. Again, an object may be in two places at one time as long as it is also in all of the places in between. When put into the language of temporal parts, this situation consists in a *temporal* part of a spatial part of that object being in one *spatial* location and a *temporal* part of another spatial part of that object being in another *spatial* location.

It appears, then, that one who employs the notion of temporal parts is committed to the position that time is in some sense prior to space and that temporal parts are the fundamental constituents of the world. If this is what is meant by the dictum "Space is in time," then these people are committed to that dictum.

But if this relation of priority holds between space and time, then there is a radical dissimilarity between space and time. Few significant analogies

exist between space and time unless the notion of temporal parts is employed. But if that notion is employed, then a difference (represented by the dictum that space is in time) exists between space and time which is so great that it cannot be offset by the lesser incomplete analogies then found to exist between space and time. So in neither case is time similar to space to any significant extent.

As the quotations at the beginning of this paper show, the holders of the manifold theory of time apparently must make use of the notion of temporal parts in stating that theory. In so far as they do make use of this notion, they are committed to the existence of the radical dissimilarity between space and time just noted. But since the manifold theory of time does not mention or represent this dissimilarity, that is, does not treat time as "prior" to space, this theory of time is to that extent not satisfactory.

University of Michigan

VIII. GOODMAN'S PARADOX AND THE PROBLEM OF RULES OF ACCEPTANCE

HOWARD SMOKLER

GOODMAN'S introduction of predicates like "grue" and "bleen" into the theory of inductive logic has occasioned a baffled and even hostile response.¹ The paradoxical consequences to which their introduction leads have impressed many people less than has the "pathological" character of these predicates. An analogy to the famous Russell paradox in the theory of sets is instructive. There too, many persons believed that the sets (or predicates) employed were "pathological" and that such sets must be banned from the language of set-theory. But the history of the problem reveals that various approaches were taken to the solution of that paradox, and that only one of these approaches involved such a ban. In the case of Goodman's paradox, there has been a similar development of alternative approaches. In this paper I want to review several of these approaches and to indicate which one seems to me to be the best. For once again, the history of the problem of the paradoxes of set-theory reveals that there is no one correct solution.

Let us first examine the paradox in the form given it by Scheffler, since his formulation reveals clearly some of its presuppositions.² For our purposes, the following ones are crucial:

(P1) Inductive rules function as rules of acceptance permitting one to assert conclusions separately.³

(P2) Any molecular predicate which is a truth-function of meaningful atomic predicates is admis-

sible in the language to which rules of inductive logic are applicable.

The inductive rule of inference to which Scheffler appeals is what he calls the "generalization formula." He describes it in the following terms:

What leads us to make one particular prediction rather than its opposite is not its deducibility from evidence, but rather its congruence with a generalization thoroughly in accord with all such evidence and the correlative disconfirmation of the contrary generalization by the same evidence. (I shall refer to this hereafter as the "generalization formula.") Of course if no relevant evidence is available to decide between a given generalization and its contrary, or if the available evidence is mixed, neither generalization will support a particular inductive conclusion.⁴

All evidence for a generalization may be supporting, or some may be supporting and some not, or none may be supporting. Only in the first case does the generalization formula (which we abbreviate as GA) apply.

Let $Ex = x$ is a specimen of emerald

$Gx = x$ is green

$Kx = x$ is examined at a time t prior to K

Consider a set of singular observation statements and consider them as all the available evidence.

1. Evidence (\mathcal{Z}) = $Ea_1 \& Ga_1 \& Ka_1$
 $Ea_2 \& Ga_2 \& Ka_2$

$Ea_n \& Ga_n \& Ka_n$

¹ A review of the history of the problem is provided in H. Kyburg's article, "Recent Work in Inductive Logic," *American Philosophical Quarterly*, vol. 1 (1964), pp. 1-39.

² I. Scheffler, "Inductive Inference: A New Approach," *Science*, vol. 127 (1958), pp. 177-181.

³ Carnap has characterized very well the notion of a rule of acceptance in the following quotation:

"It seems to me that the views of almost all writers on induction in the past and including the great majority of contemporary writers contain one basic mistake. They regard inductive reasoning as an *inference* leading from some known propositions, called the premises or evidence, to a new proposition, called the conclusion, usually a law or a singular prediction. From this point of view the result of any particular inductive reasoning is the *acceptance* of a new proposition (or its rejection, or its suspension until further evidence is found, as the case may be)."

"The Aim of Inductive Logic" in *Logic, Methodology, and Philosophy of Science*, ed. E. Nagel, P. Suppes, and A. Tarski (Stanford, Stanford University Press, 1962), pp. 316-317. In particular, an inductive rule (like GA) may be taken to function as a rule of detachment in analogy to the rule of *modus ponens* in deductive logic.

⁴ I. Scheffler, *op. cit.*, p. 177. [See n. 5 below for a less ambiguous formulation of this rule.]

2. All specimens of emerald are green. (1); GA (5)
(x) ($E_x \supset G_x$)
3. All specimens of emerald are either examined at t prior to K and are green or are not examined at t prior to K and are not green. (1); GA (5)⁵
(x) [$E_x \supset ((G_x \& K_x) \vee (\sim G_x \& \sim K_x))$]
4. a_{n+1} is a specimen of emerald and a_{n+1} is not examined at t prior to K . Hypothesis
 $Ea_{n+1} \& \sim Ka_{n+1}$
5. a_{n+1} is green. (2), (4); Universal Instantiation, M.P., Simplification
 Ga_{n+1}
6. Either a_{n+1} which is a specimen of emerald is examined at t prior to K and is green or it is not examined at t prior to K and is not green. (3); Universal Instantiation
 $Ea_{n+1} \supset [(Ga_{n+1} \& Ka_{n+1}) \vee (\sim Ga_{n+1} \& \sim Ka_{n+1})]$
7. a_{n+1} is green and is examined at t prior to K or a_{n+1} is not green and is not examined at t prior to K . (4), (6); Simplification, M.P.
 $(Ga_{n+1} \& Ka_{n+1}) \vee (\sim Ga_{n+1} \& \sim Ka_{n+1})$
8. a_{n+1} is not green and is not examined at t prior to K . (4), (7); Simplification, Truth-function
 $\sim Ga_{n+1} \& \sim Ka_{n+1}$
9. a_{n+1} is not green. Simplification
 $\sim Ga_{n+1}$

Therefore, given the same total evidence, the rules of deduction, and GA, we can inductively infer two contradictory statements. This is the paradox in the form given it by Scheffler. By an inessential modification of the argument we can arrive at Goodman's paradoxical predicates. Consider the complex statement-form

$$(A) (Gx \& Kx) \vee (\sim Gx \& \sim Kx)$$

It is of the form

$$(B) (P_x \& P_x) \vee (\sim P_x \& \sim P_x)$$

Now substitute for $\sim P_x$ the contrary of P_x , i.e., $\text{Con}(P_x)$. Then the new complex statement-form is of the form

$$(C) (P_x \& P_x) \vee (\text{Con}(P_x) \& \sim P_x)$$

Substitute for " P_x ", " x is green"; for " $\text{Con}(P_x)$ ", " x is examined at t prior to K "; and for " $\text{Con}(P_x)x$ ", " x is blue."

$$(D) (\text{Green}(x) \& Kx) \vee (\text{Blue}(x) \& \sim Kx)$$

This is the way that Goodman defines the predicate " $\text{grue}(x)$."

Most of the solutions to the paradox have resulted from a denial (or modification) of the two presuppositions of Scheffler's stated earlier in this paper. Logicians have either (a) attempted to find suitable restrictions which bar predicates of the type of " grue " and " bleen " from the language, (b) restricted the rules of inductive inference, or (c) provided a different conception of the methodology of inductive logic.

(a) The first course consists in a repudiation of (P2). A number of writers have put forth criteria which would bar predicates like " grue " and " bleen " from the language in which inductive inferences are made. The three most well-known attempts to do this are those of Carnap, of Barker

⁵ The rule GA is loose in its phrasing. If it is reformulated in the way which will be suggested the inference from (1) to (3) by means of GA' will be seen more clearly. At the same time the rule will also justify the inference from (1) to (2).

Assume the notion of property developed by von Wright in *A Treatise on Introduction and Probability* (London, Routledge and Kegan Paul, 1951), pp. 37-46. In particular, assume the notion of an implication (conditional)-property C which has the form $(K_i \supset K_j)$. Define the notion of a positive instance of a conditional-property as von Wright does in *The Logical Problem of Induction* (Oxford, Blackwell, 1957), p. 64.

(A) An individual a is a positive instance of a conditional-property $C =_{\text{df}} a$ is an individual of which C is truly predicated. Next define the contrary of a conditional-property C as follows:

(B) $\text{Con}(C) =_{\text{df}}$ if C is of the form $(K_i \supset K_j)$ then $\text{Con}(C)$ is of the form $(K_i \supset \sim K_j)$.

Now we state GA':

(C) If from $E = E_1 \& E_2 \dots \& E_n$, the total evidence available about a number of individuals a_1, a_2, \dots, a_n , it can be deduced that all of the individuals are positive instances of a conditional-property C , and from no conjunction of E_i about the individuals a_1, a_2, \dots, a_n can it be deduced that a_1, a_2, \dots, a_n are positive instances of $\text{Con}(C)$; infer $(x)(C(x))$.

That (2) follows from (1) in accordance with GA' is clear. For from (1) we can deduce that a_1, \dots, a_n are positive instances of $C = E \supset G$. But what about the inference of (3) from (1) by means of GA'? This can be shown too. From (1) it can be shown for a_i that

(D) $E \supset (G \& K)$ is truly predicable of a_i . But if (D) is the case, then so is (E).

(E) $E \supset ((G \& K) \vee (\sim G \& \sim EK))$ is truly predicable of a_i . For even though $(\sim G \& \sim K)$ is not truly predicable of a_i , the conditional property stated in (D) is truly predicable of a_i . And this is true of all the $a_i, 1 \leq i \leq n$. Therefore, by GA'

(F) $(x)(E_x \supset ((G_x \& K_x) \vee (\sim G_x \& \sim K_x)))$.

and Achinstein, and of Salmon.⁶ I believe that Goodman has dealt adequately with the first two criteria.⁷ But Salmon's attempt is an independent one. He proposes to bar from any language in which the straight rule of induction is applicable to arguments all those terms which are basic but not purely ostensive. Intuitively, he thinks that "grue" and "bleen" are putatively ostensive predicates but that they are really not acceptable ostensive predicates. A predicate is a purely ostensive predicate if and only if:

- (1) It can be defined ostensively.
- (2) Its positive and negative instances for ostensive definition can be indicated non-verbally.
- (3) The respects in which the positive instances resemble each other and differ from the negative instances are open to direct inspection, i.e., the resemblance in question is an observable resemblance.⁸

Salmon claims that predicates like "grue" and "bleen" fulfill conditions (1) and (2) but that they fail to fulfill condition (3). What is in question is the meaning of the relation "observable resemblance." For his argument to be valid, Salmon must assume that: (i) the relation is independent of the language in which the terms denoting the properties appear and (ii) that it is a relation which is logically weaker than "matches" but stronger than "is similar to." The first condition would have to be satisfied for the argument to be valid. If it were not satisfied, then it might be that in a language in which "grue" and "bleen" were terms which we learned at our mother's knee, positive instances of "grue" or "bleen" would observably resemble each other. The second condition would have to be satisfied, for as Goodman has pointed out, the relation of "matching" is too strong even to hold of all cases of green while the relation of "is similar to" holds of all individuals. But *a priori* there is no reason to believe that "is an observable resemblance" fulfills both of these conditions. Or

put in other words, in a "grue-bleen" language, "grue" and "bleen" might well be purely ostensive predicates while "green" and "blue" would not. And there seems to be no reason, in principle, to choose the one language rather than the other.

An examination of all these attempts, that of Carnap and of Barker-Achinstein as well as that of Salmon, reveals that at bottom the same basic criticism can be made of all of them. It is this. If we consciously discount the factor of habitual linguistic usage, all the criteria so far devised are perfectly symmetrical as to the two sets of concepts, "grue-bleen" and "green-blue." Obviously, no criterion which bans predicates can be formulated on the basis of such a factor. For whose habits are to be considered? And how unhabitual must the predicate be to be banned? In fact, it seems unlikely that any devisable criterion of this type will be capable of avoiding some version of this criticism. That the two sets of terms are interdefinable is not subject to question. Therefore, a criterion must distinguish the two sets of terms on the basis of some semantic relationship that one set possesses and the other does not. But such a relationship would have to have the features enunciated above, and I do not believe that many philosophers of an empiricist bent would seriously look for such a relationship.

(b) Some logicians have taken another approach; they have reformulated the principles of inductive logic so as to avoid the paradox. In fact, their work can be viewed as a denial (in a qualified sense) of P₂. This is the path taken by Goodman. In effect, he has taken explicit account of the factor of habitual linguistic usage in his modification of the traditional rules of induction. But he does accept P₁, a rule of acceptance. Since new predicates of this paradoxical type are not systematically eliminable from the language in which inferences are formulable, some way must be found of disqualifying at least one of the generalizations from which a singular statement is derivable. In

⁶ R. Carnap, "On the Application of Inductive Logic," *Philosophy and Phenomenological Research*, vol. 8 (1947-48), pp. 133-148; see especially p. 146; S. Barker and P. Achinstein, "On the New Riddle of Induction," *Philosophical Review*, vol. 69 (1960), pp. 511-522; W. Salmon, "On Vindicating Induction," *Philosophy of Science*, vol. 30 (1963), pp. 252-261; see especially pp. 259-261.

⁷ Goodman has answered Carnap in *Fact, Fiction, and Forecast* (Cambridge, Harvard University Press, 1955), especially pp. 78-79. He has answered Barker and Achinstein in his article, "Positionality and Pictures," *Philosophical Review*, vol. 69 (1960), pp. 523-525.

⁸ W. Salmon, *op. cit.*, p. 259. According to Salmon:

An ostensive definition contains three parts: 1. The indication of a number of positive instances, i.e., individuals which have the property to be denoted by that constant. 2. The indication of a number of negative instances, i.e., individuals which lack the property to be denoted by that constant. 3. A similarity clause stating that anything resembling all the positive instances in some respect, provided it does not also resemble any of the negative instances in that respect, also has the property to be denoted by the instance.

particular, in the derivation of the paradox given above, the inference to (3) must be blocked. Goodman does this by introducing the notion of "entrenchment" and by providing rules which, under certain conditions, eliminate statements containing less entrenched predicates. In a certain sense, Goodman's course is that of eliminating certain predicates from the language. But of course there are no mechanical rules for distinguishing better from less entrenched predicates; this varies from case to case.

In a given context, one predicate may be much better, equally, or much less entrenched than another. To determine this we make use of available linguistic information. A term which has appeared in more projected, unexhausted, and unviolated hypotheses than has another term is better entrenched than the second term. Intuitively we know (except for borderline cases) when one term is better entrenched than another. But Goodman's notion is subject to criticism. The choice of "better entrenched" predicates as a basis for inference is logically arbitrary. There is no reason why we should not choose "worsely entrenched" predicates, for in this case too we would avoid the paradox. Goodman might answer that the choice in fact reflects our intuitions of what are correct inferences. But when it is exactly our intuitions which are called into question by the existence of the paradox, this does not seem a sufficient answer to the problem.

This arbitrariness at the basis of Goodman's positive system is, I believe, the reason that leads many logicians to be wary of it. As Kyburg points out, Goodman must assume some variant of GA to justify the use of "better entrenched" predicates.⁹ But once more we can point to the fact that it is this very principle of enumerative induction which is called into question by Goodman's paradox. Therefore it is not justifiable to use it as an assumption.

Is there any way out of the paradox which Goodman has posed? Or must we abandon the attempt to formulate an inductive logic? In my opinion, there is a way out, and the way has already been taken by Carnap. It involves (c) providing a different conception of the methodology of inductive logic. More specifically it

involves rejecting (P₁), the employment of rules of acceptance in inductive logic. That the employment of rules of acceptance leads to contradiction has been specifically recognized by Bar-Hillel, but I do not believe the point has been sufficiently stressed.¹⁰

First we must note how the refusal to allow inductive rules to be treated as rules of acceptance avoids the paradox. If Scheffler's argument is reformulated in these terms, two statements may be asserted:

(A) [(1) & (4)] confirm (5)

(B) [(1) & (4)] confirm (9)

and these two statements do not contradict each other nor are they even the contraries of one another. Scheffler employs a qualitative notion of confirmation. If a quantitative notion is employed, then statements of the following kind would be introduced:

(A') Conf [(5) | (1) & (4)] = m where m and n

(B') Conf [(5) | (1) & (4)] = n are real numbers in the interval (0, 1)

Surely these two statements do not contradict one another, especially if $m \neq n$. But Leblanc has argued on plausible grounds that m must equal n and that in particular $m = n = 0.5$.¹¹ His assumptions are two in number; (1) that given evidence statements of the form

$[A(a_1) \& E(a_1) \& B(a_1)] \& [A(a_2) \& E(a_2) \& B(a_2)] \& \dots [A(a_k) \& E(a_k) \& B(a_k)]$

and

$A(b) \& \sim E(b)$

that for any predicate "B," the degree of confirmation of

$B(b)$

is the same, and (2) that logically equivalent statements can replace one another (given certain conditions) in confirmation statements as well as in a form of the theorem of total probabilities. Assumption (2) does not seem open to question, but assumption (1) does. As Leblanc himself admits, Carnap in his system of confirmation rejects it. And Leblanc may be said to have shown the implausibility of assumption (1). But even if, for the sake of argument, both assumptions are granted, the conclusions are only somewhat

⁹ H. Kyburg, *op. cit.*, p. 18.

¹⁰ "Discussion of Salmon's Paper" in *Induction, Some Current Issues*, ed. by H. Kyburg and E. Nagel (Middletown, Wesleyan University Press, 1963), p. 46.

¹¹ H. Leblanc, "That Positive Instances Are No Help," *Journal of Philosophy*, vol. 60 (1963), pp. 453-462; see especially pp. 461-462.

counter-intuitive; they do not lead to a contradiction.

Yet the following argument can be offered against the approach sketched here.¹² Carnap's methodology of inductive logic involves the interpretation of degrees of confirmation as fair betting quotients.¹³ It is perfectly true that (A') is neither the contradictory nor the contrary of (B'). But what fair bet on (5) can be made given (1) & (4)? And what fair bet on (9) can be made given (1) and (4)? The interpretation provided by Carnap of the fair betting quotient (which we call the first condition) guarantees that the sum of the odds given on (5) and on (9) must equal 1.¹⁴ Thus the following assignment of bets, which is intuitively acceptable, can be made in accordance with this criterion:

(A'') On (5), odds of 10 : 1 that (5) is true

(B'') On (9), odds of 1 : 10 that (9) is true

There is another intuition (which we call the second condition) that recommends itself very strongly. It is this. We should bet (at much more than even odds) on the occurrence of a property which has always been exemplified in the total evidence available to us. In this case that property is the complex property

$(Gx \ \& \ Kx) \vee (\sim Gx \ \& \ \sim Kx)$.

This principle recommends to us an assignment much like the following one:

(A''') On (5), odds of 10 : 1 that (5) is true

(B''') On (9), odds of 8 : 1 that (9) is true

It is clear that no assignment of odds can fulfill both the first and second conditions. Therefore no bets can be placed on these propositions. As a consequence, Carnap's interpretation of confirmation fails in this case, and no solution has been provided to the paradox.

I believe that the difficulty posed is a serious but not a fatal one. Formally, of course, the problem is easily disposed of. All that is required is that the bettor bet in accordance with a logical probability function. Let him reject the second condition; indeed, let him bet in some such way as $(A'') - (B'')$. But this answer, although formally unobjectionable, raises more questions than it solves. Why should the bettor abandon an intuition which seems perfectly reasonable? Are we not abandoning the search for an adequate confirmation function defined for a language of a specified structure, a task which is the goal of inductive logic? And would not the hypothetical bettor whom we have described, if pushed to disclose his reason for ignoring this intuition, be forced to admit to a criterion for selection of odds very much like the notion of entrenchment?¹⁵

These questions suggest that the answer so far given is not compatible with the approach I have sketched. But I believe that there is a way out of these difficulties. For the language system implicitly employed by Scheffler and Goodman, a first-order functional calculus, the second condition seems reasonable. But Goodman and Scheffler have not chosen the proper language in which to express these predicates. All the indications are that a language which Carnap calls a "coordinate language" is a more appropriate one.¹⁶ Some notion of order is required to characterize adequately such predicates as "grue" or "bleen." For such languages the intuitions described above in the second condition do not hold.¹⁷ Investigation of the inductive properties of such languages has only begun.¹⁸ There is reason to believe that in them an adequate confirmation function can be defined which is a fair-betting function and which does not violate our intuitions. Indeed in such a

¹² I wish to thank the referee for bringing this point to my attention.

¹³ Carnap states this explicitly in the article, "Remarks on Probability," *Philosophical Studies*, vol. 14 (1963), pp. 65-75; see especially p. 67.

¹⁴ This follows from a theorem of confirmation theory: if e is not logically false then $c(h, e) + c(\sim h, e) = 1$, and from the fact that a function is a fair betting quotient if and only if it obeys the axioms of confirmation theory. For a proof of the latter fact, see J. Kemeny, "Fair Bets and Inductive Probabilities," *Journal of Symbolic Logic*, vol. 20 (1955), pp. 263-273.

¹⁵ It is true that a person might object to betting with such high odds on a peculiar predicate of the type specified above. But why? This once more raises the whole problem previously discussed by Goodman.

¹⁶ Carnap has described the character of such a language in various places. See *Logical Foundations of Probability* (Chicago, University of Chicago Press, 1962), pp. 64-65, and *Introduction to Symbolic Logic and its Applications* (New York, Dover Press, 1958), pp. 161-171.

¹⁷ See P. Achinstein's investigation, "Confirmation Theory, Order, and Periodicity," *Philosophy of Science*, vol. 30 (1963), pp. 17-35, and R. Carnap's reply, "Discussion: Variety, Analogy, and Periodicity in Inductive Logic," *Philosophy of Science*, vol. 30 (1963), pp. 222-227.

¹⁸ See P. Achinstein, *ibid.*, and the article of R. Carnap cited in n. 17. See also H. Putnam, "Degree of Confirmation and Inductive Logic" in *The Philosophy of Rudolf Carnap* ed. P. Schilpp (LaSalle, Ill., Open Court, 1963), pp. 761-784, as well as Carnap's reply to Putnam, "Replies and Systematic Expositions," in the same book, pp. 983-989, sect. 29.

language there is reason to believe that both "grue-bleen" and "green-blue" are acceptable predicate terms. But there is no reason to think that their employment will lead to paradox.

Carnap's formulation of inductive logic avoids the paradox in its crippling form, and this is surely one reason to accept it. But more generally, rules of acceptance are in serious trouble. In a variety of systems, rules of acceptance have been shown to lead either to other paradoxes than the Goodman one or to lead to counterintuitive conclusions. Several of these results should be listed:

(A) Schick has shown that a restricted rule of detachment in Kyburg's system leads to inconsistency in the system.¹⁹

(B) Hempel's rule of acceptance, if combined with his criterion for rational belief, leads to a variant of the lottery paradox. This has been shown by Kyburg.²⁰

(C) Levi has shown that a non-probabilistic notion of confirmation like Popper's, if combined with a rule of acceptance, leads to intuitively unacceptable conclusions.²¹

In the light of these results, Goodman's paradox should be considered as an independent argument against a conception of inductive logic which makes use of rules of acceptance. There are independent arguments for and against such a view of inductive principles.²² Goodman's paradox must surely be counted an argument for renouncing them.

Stanford University

¹⁹ F. Schick, "Consistency and Rationality," *Journal of Philosophy*, vol. 60 (1963), pp. 5-19. See also Kyburg, "A Further Note on Rationality and Consistency," *Journal of Philosophy*, vol. 60 (1963), pp. 463-465.

²⁰ Hempel's rule of acceptance for inductive logic as well as his criteria for rational belief are to be found in his article "Deductive-Nomological vs. Statistical Explanation" in vol. III of *Minnesota Studies in the Philosophy of Science*, ed. by H. Feigl and G. Maxwell (Minneapolis, University of Minnesota Press, 1962), pp. 98-169. See especially sect. 12. Kyburg's demonstration is contained in his paper, "Probability, Rationality, and a Rule of Detachment," to be published in *Proceedings of the 1964 International Congress of Logic, Methodology, and Philosophy of Science*. Kyburg supports the search for rules of detachment.

²¹ I. Levi, "Corroboration and the Rule of Acceptance," *British Journal for the Philosophy of Science*, vol. 13 (1963), pp. 307-313.

²² For arguments against rules of acceptance in inductive logic see Carnap (n. 3). See also R. Jeffrey, "Valuation and Acceptance of Scientific Hypotheses," *Philosophy of Science*, vol. 33 (1956), pp. 237-246. For arguments in favor of the necessity of rules of acceptance see H. Putnam, *op. cit.*, as well as the references in n's. 10 and 20.

IX. THREE KINDS OF CLASSES

HUGH S. CHANDLER

THE difficulties in which we find ourselves when we struggle with the notion that all classes¹ are "artificial," or become entangled in the debate between nominalists and realists,² as well as many other difficulties, arise in part from our lack of reasonably precise technical terms with which to discuss classes. In this paper I want to help remedy this situation. I am going to describe three quite different kinds of classes of things. These were suggested to me by certain remarks made by William Whewell and Dugald Stewart; but my interest here is not at all scholarly. I shall feel free to lay down definitions that may not correspond to what Whewell, or Stewart, had in mind.

I. CLOSED CLASSES

Whewell said: "A Natural Group is steadily fixed, though not precisely limited; it is given in position, though not circumscribed; it is determined, not by a boundary without, but by a central point within; — not by what it strictly excludes, but by what it eminently includes; — by a Type, not by a Definition."³ Perhaps the contrast being sketched here between classes determined by definition and classes determined by type is the contrast between what I call "closed classes" and "type-governed classes."

A class of things, *K*s, is a closed class if and only if each and every imaginable *K* has all of the distinguishing features of *K*s.

Here is what I mean by a "distinguishing feature." If there are two imaginable things, *x* and *y*, such that *x* has a certain feature, *A*, and *y* does not have that feature, and this difference is what makes *x* a *K* and *y* not a *K*, then *A* is a distinguishing feature of *K*s.

When I say that some feature, *A*, makes a thing, *x*, a *K* and the absence of this feature in another thing, *y*, makes *y* not a *K*, I do not mean that *A* is

a necessary or a sufficient condition for somethings being a *K*. I mean only that *in regard to these two things* it is the presence or absence of feature *A* that is making the difference between being a *K* and not being a *K*. For example I can imagine two artifacts, *x* and *y*, such that *x* is a bottle and *y* is not and such that it is *x*'s narrow neck that makes the difference. Thus having a narrow neck is a distinguishing feature of bottles. Obviously having a narrow neck is neither a necessary nor a sufficient condition for being a bottle. It should also be noted that, speaking abstractly, it may not be only its lack of a narrow neck that makes *y* not a bottle, although it is certainly not only its narrow neck that makes *x* a bottle; but when we are contrasting these two artifacts it is not only proper but correct to say that it is simply *x*'s narrow neck and *y*'s lack of such a neck that is making the difference between being a bottle and not being a bottle. But for its lack of a narrow neck *y* would be a bottle like *x*.

All imaginable geometrical triangles have all of the distinguishing features of geometrical triangles. (E.g., all imaginable geometrical triangles are closed figures, and this is a distinguishing feature of geometrical triangles.) Closed classes are at home in mathematics and formal logic. We certainly must not assume that most classes are of this sort.

Many closed classes are classes that violate what might be called Kant's law of *logical continuity*. Kant expresses the law like this:

... there are no species or sub-species which [in the view of reason] are the nearest possible to each other; intermediate species or sub-species being always possible, the difference of which from each of the former is always smaller than the difference existing between these.⁴

Hamilton puts it like this:

... no two coördinate species touch so closely on each

¹ The term "class" is not being used in any technical sense. This is the term used in the discussion of my topic by Whewell and John Stuart Mill. I use it as they did.

² One aspect of this debate is considered by way of a conclusion to this paper.

³ William Whewell, *The Philosophy of the Inductive Sciences*, vol. I (London, John W. Parker, 1840), p. xxxiii.

⁴ Immanuel Kant, *Critique of Pure Reason*, tr. J. M. D. Meiklejohn (London, J. M. Dent & Sons, 1945), p. 382.

other, but that we can conceive other or others intermediate.⁵

In modern terms we might say "every class is bounded, if not by actual then at least by imaginable, borderline cases." Or we might say "all terms are vague." When modern philosophers say that a term is "vague" they do not always mean that there is something wrong with the term. Sometimes they mean only that the term conforms to the law of logical continuity.⁶ As Kant and Russell saw, this law implies that every class is virtually bounded by an infinity of classes.⁷

The law does not apply to many of the classes of mathematical and logical entities. Hamilton rightly says:

... it breaks down when we apply it to mathematical classifications. Thus all angles are either acute or right or obtuse. For between these three coördinate species or genera no others can possibly be interjected, though we may always subdivide each of these, in various manners, into a multitude of lower species.⁸

Since most closed classes are classes of such entities it is clear that many closed classes violate Kant's law.

II. TYPE-GOVERNED CLASSES

We can define a type-governed class like this: A class of things, *Ks*, is a type-governed class if and only if there are imaginable *Ks* that have all of the distinguishing features of *Ks*, and there are imaginable *Ks* that lack one or more of the distinguishing features of *Ks*.

The type, or paradigm, of a type-governed class, *Ks*, is a *K* that has all of the distinguishing features of *Ks*. It may be helpful to picture the type as being at the "center" of the class. Whewell said:

A Type is an example of any class, for instance a species of a genus, which is considered as eminently possessing the character of the class. All the species which have a greater affinity with this type-species than with any others, form the genus, and are ranged about it, deviating from it in various directions and different degrees. Thus a genus may consist of several species which approach very near the type, and of

which the claim to a place with it is obvious; while there may be other species which straggle further from this central knot, and which yet are clearly more connected with it than with any other. And even if there should be some species of which the place is dubious, and which appear to be equally bound to two generic types, it is easily seen that this would not destroy the reality of the generic groups, any more than the scattered trees of the intervening plain prevent our speaking intelligibly of the distinct forests of two separate hills.⁹

One common interpretation of what Wittgenstein meant by a "family resemblance" among the members of a class seems to be that he meant that the class was a type-governed class. Thus one philosopher says:

It is often admitted, in the analytical treatment of some fairly specific concept, that the wish to understand is less likely to be served by the search for a single strict statement of the necessary and sufficient conditions of its application than by seeing its applications—in Wittgenstein's simile—as forming a family, the members of which may, perhaps, be grouped around a central paradigm case and linked with the latter by various direct or indirect links of logical connexion and analogy.¹⁰

And another says:

It should be carefully noted that only fully developed mystical experiences are necessarily apprehensive of the One. Many experiences have been recorded which lack this central feature but yet possess other mystical characteristics. These are borderline cases, which may be said to shade off from the central core of cases. They have to the central core the relation which some philosophers like to call "family resemblance."¹¹

These statements are not clear; but the imagery is imagery that has been used in describing a type-governed class.

III. INDETERMINATE CLASSES

I believe that in his *Philosophical Essays* Dugald Stewart describes a sort of term that does not

⁵ William Hamilton, *Lectures on Logic* (Boston, Gould and Lincoln, 1860), p. 149.

⁶ See, for example, H. G. Wells, *First and Last Things* (London, Watts & Co., 1929), pp. 13-16.

⁷ Kant, *op. cit.*, p. 383. Bertrand Russell, "Vagueness," *The Australasian Journal of Philosophy*, vol. 1 (1923).

⁸ Hamilton, *op. cit.*, p. 149.

⁹ Whewell, *op. cit.*, pp. 476-477.

¹⁰ P. R. Strawson, *Individuals* (London, Methuen & Co., 1961), p. 11.

¹¹ Walter T. Stace, *The Teachings of the Mystics* (New York, Mentor Books, 1960), p. 15.

denote either things forming a closed class or things forming a type-governed class.

I shall begin with supposing, that the letters *A, B, C, D, E*, denote a series of objects; that *A* possesses some one quality in common with *B*; *B* a quality in common with *C*; *C* a quality in common with *D*; *D* a quality in common with *E*;—while, at the same time, no quality can be found which belongs in common to any *three* objects in the series. Is it not conceivable, that the affinity between *A* and *B* may produce a transference of the name of the first to the second; and that, in consequence of the other affinities which connect the remaining objects together, the same name may pass in succession from *B* to *C*; from *C* to *D*; and from *D* to *E*? In this manner, a common appellation will arise between *A* and *E*, although the two objects may, in their nature and properties, be so widely distant from each other, that no stretch of imagination can conceive how the thoughts were led from the former to the latter.¹²

Stewart held that beautiful things form such a class.¹³

I picture an indeterminate class a bit differently. Let us assume that *felonies* may be taken as an example.¹⁴ There are different kinds of felonies, and these are distinguished from non-felonies, i.e., from other kinds of offence, by different distinguishing features. I assign Roman numerals to the different kinds of felony, and Arabic numerals to the distinguishing features. Thus the distinguishing features of felonies of kind I are features 1, 2, 3, and 4. Here, then, are some kinds of felonies and their respective distinguishing features:

<i>Felonies</i>	<i>Features</i>
I	1234
II	2345
III	3456
IV	4567
V	5678
VI	6789
etc.	etc.

The first kind of felony has four distinguishing features. The second "resembles" the first in that it has three of these features, but it lacks one of the features of the first kind, and, in addition it has a distinguishing feature that the first kind does not have. And so it goes from one sort of felony to the

next, with distinguishing features appearing and dropping out until we find kinds of felonies with no distinguishing features at all in common. The similarity between this imagery and the imagery Wittgenstein uses in describing a "family resemblance" is to be noted.¹⁵ My guess is that Wittgenstein would have spoken of the members of an indeterminate class, as well as of the members of a type-governed class, as bearing only a "family resemblance" to each other.

Of course there is no reason for us to stop at six or seven kinds of felony. The imagery suggests that we can perfectly well have some kind of felony with distinguishing features 8, 9, 10, and 11. There is no limit drawn here to the number of different kinds of felony there may be.

In this imagery we have no type, or paradigm. We have not mentioned that kind of felony that has all of the distinguishing features of felonies, and the imagery, as it were, provides no room for such a felony. The imagery suggests that for any kind of felony we choose, *x*, there may be another kind of felony, *y*, such that *x* has no distinguishing feature in common with *y*.

There are, then, differences between the imagery here and the imagery we used in discussing type-governed classes. But if there is to be a useful distinction between type-governed classes and indeterminate classes, or, for that matter, between closed classes and indeterminate classes, we must find differences other than mere differences in imagery. When I say that the class of things, *Ks*, is an indeterminate class, I shall mean that the distinguishing features of *Ks* form two or more clusters, *C*₁, *C*₂, *C*₃, . . . etc., such that one or more of the distinguishing features in *C*₁ are also distinguishing features in *C*₂, and one or more of the distinguishing features in *C*₂ are also distinguishing features in *C*₃, etc., and such that it is impossible for one and the same *K* to have all of the distinguishing features in any pair of clusters. It is to be noted that the three kinds of classes I have now defined are such that it is impossible for any one class to be more than one of these kinds. A class of things, *Ks*, cannot, for example, be both a closed class and an indeterminate class. Furthermore, since I include all *imaginable Ks* in the class of *Ks* it is impossible for the members of one kind of

¹² Dugald Stewart, *The Works of Dugald Stewart*, vol. IV (Cambridge, Hilliard and Brown, 1829), p. 187.

¹³ *Ibid.*, pp. 195-264.

¹⁴ This is the example given by John Stuart Mill, *A System of Logic* (New York, Harper & Bros., 1846), p. 25, n. Mill also was inclined (cf. p. 408) to agree with Stewart that the class of beautiful things is an indeterminate class.

¹⁵ See, for example, *Philosophical Investigations*, §§ 66, 67.

class to be perfectly co-extensive with the members of one of the other two kinds of classes.

According to my definition, if felonies form an indeterminate class, it must be impossible for there to be a kind of offence that has all of the distinguishing features of felonies. Such an offence is impossible because at least two of the distinguishing features of felonies are incompatible.

It does not follow from the fact that there are two kinds of *K* which are distinguished from each other by incompatible distinguishing features that *K*s form an indeterminate class. The distinguishing features of an equilateral triangle in geometry are incompatible with the distinguishing features of a right-angled triangle. But it does not follow that geometrical triangles form an indeterminate class. It is perfectly possible for a triangle to have all of the distinguishing features of geometrical triangles. Thus if felonies form an indeterminate class, it is not because there are different kinds of felonies distinguished from each other by features such that no offence could have all of them; rather it is because the features that distinguish felonies from non-felonies are such that no one offence could have all of them.

In the case of geometrical triangles we might say that over and above the diverse features that distinguish the various kinds of triangles from each other there is the set of distinguishing features of geometrical triangles, and these features are such that all geometrical triangles have all of them. We might picture the situation like this: Let features *a*, *b*, and *c* be the distinguishing features of geometrical triangles, and let features 1, 2, 3, 4, etc., be the distinguishing features of the various kinds of geometrical triangles. Then the various triangles might be thought of like this:

<i>Triangle</i>		
I	<i>a b c</i>	1 2 3
II	<i>a b c</i>	2 3 4
III	<i>a b c</i>	3 4 5
IV	<i>a b c</i>	4 5 6
etc.	etc.	etc.

Where there is a type-governed class, as where there is a closed, there may be incompatible kinds. Suppose, for example, that the class of dicotyledons is a type-governed class.¹⁶ In that case there will be dicotyledons that have all of the distinguishing features of dicotyledons, and there will be dicotyledons that lack one or more of the distinguishing

features of dicotyledons. But, in such a case, there is nothing to prevent some of the various kinds of dicotyledon from having incompatible distinguishing features. If features *a*, *b*, *c*, and *d*, are the distinguishing features of dicotyledons, and features 1, 2, 3, 4, etc., are the distinguishing features of the various kinds of dicotyledon, we can picture the situation like this:

<i>Dicotyledon</i>		
I	<i>a b c d</i>	1 2 3
II	<i>a b c</i>	2 3 4
III	<i>a b d</i>	3 4 5
IV	<i>b c d</i>	4 5 6
etc.	etc.	etc.

Here it is not the case that all dicotyledons have all of the distinguishing features of dicotyledons; but it is the case that there are some dicotyledons that have them all. The incompatible features here, i.e., features 1 and 4, features 2 and 5, etc. (e.g., having leaves characteristically compound, and having leaves characteristically doubly compound) are distinguishing features of various kinds of dicotyledon.

We can set out the nature of a geometrical triangle, or of a dicotyledon; but (assuming that the class of felonies is an indeterminate class) if we are asked to set out the nature of a felony we can only describe some of the diverse natures of some of the different kinds of felony. Felonies as such do not have a nature. Thus Mill says: ". . . there is no lawyer who would undertake to tell what a felony is, otherwise than by enumerating the various kinds of offence which are so called."¹⁷ We might express this by saying that felonies have only an indefinitely large number of related natures; or, since it comes to the same thing, that the word "felony" does not have a meaning, but only an indefinitely large number of related meanings.

The class of beautiful things is, I believe, importantly different from the class of felonies in that the former crosses categories while the latter does not. A class of things, *K*s, crosses categories if and only if there are imaginable *K*s belonging to one category and there are imaginable *K*s belonging to a different category.

When I say that two things, *x* and *y*, belong to different "categories" I mean that there is a pair of contrary attributes *A* and *B* such that one or both judgments of the form "*x* is *A*" and "*x* is *B*" make sense, but such that neither a judgment of

¹⁶ See William L. Davidson, *The Logic of Definition* (London; Longmans, Green, & Co., 1885), p. 284.

¹⁷ J. S. Mill, *op. cit.*, p. 25. My italics.

the form " y is A ," nor of the form " y is B " makes sense.¹⁸

A pair of attributes, A and B , are "contrary attributes" if and only if there is some judgment of the form " x is not A " which implies that x is B , or, if A and B form a continuum, in the border area between A and B ; and there is some judgment of the form " x is not B " which implies that x is A , or, if A and B form a continuum, in the border area between A and B .

"My wife is not well" implies that the speaker's wife is sick, and "My wife is not sick" implies that the speaker's wife is well. Thus sick and well are contrary attributes. "My wife is well" and "My wife is sick" make sense; but "My vacation is [or was] well" and "My vacation is [or was] sick" do not. Thus one's wife and one's vacation are in different categories. Now there are both wives and vacations in the class of beautiful things, so this class crosses categories. It does not follow from this that beautiful things form an indeterminate class. There is no reason I can see to think that a type-governed class, or even a closed class, cannot cross categories. I believe that the class of beautiful things is an indeterminate class; but this requires a separate proof.

Let me now just hint at how all this contributes to the discussion of realism. Porphyry said: "... as to what concerns genus and species, the question is to know if they are realities subsisting in themselves, or are merely simple conceptions of the mind, . . ."¹⁹ A realist is one who answers that a genus or a species is a reality subsisting in itself. Does this view imply that all generic and specific terms indicate *closed classes*? It may well be that

Plato, Anselm, Bernard of Champeaux, and other realists, have held both of these views, and there may be a natural affinity between them; but there is no reason to suppose that either view *implies* the other. A realist, I should suppose, can hold that the classes indicated by such terms are *either* closed or type-governed, for the members of either of these two sorts of classes *share a nature* by virtue of which they are members of the class; although they do this in different ways.

The later Wittgenstein used the notion of "family resemblance" to attack the doctrine that most ordinary classes are closed classes. Wittgenstein, like Whewell and others, argued that in the case of many everyday terms, ' K ', there are K s that lack "characteristic" features of K s. But Renford Bambrough has said recently that Wittgenstein's discussion of "family resemblance" also constitutes a refutation of "realism."²⁰ It seems clear that Bambrough has not noticed the crucial difference between type-governed classes and indeterminate classes. Perhaps we strike a blow against realism if we show that the classes realists thought were informed by one single Idea are in fact indeterminate classes. But this is not what Wittgenstein has done. Wittgenstein nowhere argues that the characteristic features of most common kinds of things form incompatible clusters, i.e., make up indeterminate classes, even though his imagery is such as to hint, sometimes, that he is thinking of classes of this sort. For this reason I do not think that his crushing attack on the notion that most of the classes indicated by everyday terms are closed classes should be taken as a defeat of realism.

University of Illinois

¹⁸ This definition generates many fewer "categories" than does the one offered by Ryle. See Gilbert Ryle, "Categories" in *Logic and Language* (2nd series) ed. A. G. N. Flew (New York, Philosophical Library, 1953), pp. 65-81.

¹⁹ Quoted from Anne Fremantle, *The Age of Belief* (New York, Mentor Books, 1954), p. 20.

²⁰ Renford Bambrough, "Universals and Family Resemblances," *Proceedings of the Aristotelian Society*, vol. 61 (1961), pp. 207-222.

X. PART X OF HUME'S *DIALOGUES*

WILLIAM H. CAPITAN

IN Part X of Hume's *Dialogues Concerning Natural Religion*, Philo presents the famous trilemma attributed to Epicurus: "Is God willing to prevent evil, but not able? Is he able, but not willing? Is he both willing and able? Whence then is evil?" Some critics say Philo is trying to disprove God's existence.¹ Some say he is not.² Actually, he is demolishing natural religion, not by disproving God's existence, but by invalidating the argument to God's moral attributes. I would like to show how he does this.

Natural religion, in the *Dialogues* and commonly in the eighteenth century, is the set of beliefs about God allegedly derivable from reason and experience unaided by revelation. Natural theologians claimed to know, not only that God exists, but also enough about God's nature to infer that men ought to worship him by being pious and virtuous, that men must repent of their sins, and that there are present and future rewards and punishments.³

The position Philo takes against natural religion is, as he says, "moderate scepticism." He questions only the adequacy of the evidence offered by natural theologians for their claims about God's nature. He does not, as would a Pyrrhonist, reject the common sense idea of evidence; for he believes arguments derived from common life can dispel the subtle arguments of the sceptics. "But," he adds, "it is evident whenever arguments lose this advantage and run wide of common life, that the most refined scepticism comes to be upon a footing with them, and is able to oppose and counter-

balance them. The mind must remain in suspense between them; and it is that very suspense or balance which is the triumph of scepticism."⁴ Philo maintains this position throughout the *Dialogues*, and this is the key to his argument in Part X.

Demea opens Part II by saying: "The question is not concerning the *being* but the *nature* of God. This, I affirm, from the infirmities of human understanding, to be altogether incomprehensible and unknown to us" (Hume, p. 15). And, whatever was said earlier or will be said later, Philo agrees with him here and many times after: "the question can never be concerning the *being* but only the nature of the Deity. The former truth . . . is unquestionable and self-evident. Nothing exists without a cause; and the original cause of this universe (whatever it be) we call God . . ." (Hume, p. 16). Philo does not depart from his moderate scepticism here; for, unlike the natural theologians' abstruse reasonings about God's nature, this reasoning about God's existence rests on the solid ground of common sense.

Basson says this agreement is not meant seriously.⁵ He believes a substantial part of the *Dialogues* is concerned with the question of existence and the question of God's existence and the question of God's nature could hardly be discussed independently because the former cannot be other than a question of something's having certain antecedently specified characteristics. But Philo has already said there is an original cause of

¹ Among these are: T. H. Huxley, *Hume* (London, 1879), pp. 146-152; A. H. Basson, *David Hume* (Baltimore, 1958), pp. 105-106; Nelson Pike, "Hume on Evil," *Philosophical Review*, vol. 72 (1963), pp. 180-197, *idem* in his introduction to *God and Evil* (Englewood Cliffs, N. J., 1964).

² Among these are: N. Kemp-Smith, Introduction, *Hume's Dialogues Concerning Natural Religion* (New York, 1948), pp. 67-69; F. Copleston, *A History of Philosophy*, vol. 5 (Westminster, Md., 1959), pp. 307-309; R. J. Butler, "Natural Belief and the Enigma of Hume," *Archiv Für Geschichte der Philosophie*, Band 42/Heft 1 (1960), pp. 73-100.

³ For example, see: Charles Blount, *The Oracles of Reason* (London, 1693); Mathew Tindal, *Christianity as Old as the Creation, or the Gospel a Republication of the Religion of Nature* (London, 1730).

⁴ David Hume, *Dialogues Concerning Natural Religion*, ed. Henry Aiken (New York, 1962), p. 10. All subsequent references to this work are to this edition and will appear in the text as "Hume." I shall not ask whether Philo speaks for Hume. On this point see Copleston, *op. cit.*, pp. 308-309.

⁵ Basson, *op. cit.*, pp. 105-106.

the universe and we call it God.⁶ This is to say something exists with the attribute of being cause of the universe, an attribute commonly associated with the name of "God." So the disputants are asking whether the cause of the universe has other attributes commonly associated with the name "God." Responding to this question, then, Philo makes two assertions which determine the course of the entire discussion.

First, he gives his view of how we associate certain expressions with the name "God." "Wisdom, thought, design, knowledge—these we justly ascribe to him because these words are honorable among men, and we have no other language or other conceptions by which we can express our adoration of him. But let us beware lest we think that our ideas anyway correspond to his perfections, or that his attributes have any resemblance to these qualities among men. He is infinitely superior to our limited view and comprehension, and is more the object of worship in the temple than of disputation in the schools" (Hume, p. 16).

Second, Philo challenges Cleanthes by saying: "Our ideas reach no farther than our experience: We have no experience of divine attributes and operations: I need not conclude my syllogism: You can draw the inference yourself" (Hume, pp. 16–17). If Philo wants Cleanthes to infer that God does not exist, he should not have conceded earlier that there must be an original cause of the universe. He wants Cleanthes to infer that we cannot prove anything about God which would make natural religion seem reasonable.⁸

Cleanthes responds to Philo's challenge with the argument from design:

The curious adapting of means to ends, throughout all nature, resembles exactly, though it much exceeds,

the productions of human contrivance—of human design, thought, wisdom, and intelligence. Since therefore the effects resemble each other, we are led to infer, by all the rules of analogy, that the causes also resemble, and that the Author of nature is somewhat similar to the mind of man, though possessed of much larger faculties, proportioned to the grandeur of the work which he has executed. (Hume, p. 17.)

He wants to prove God is, not just cause of the universe, but also "similar to the mind of man"; for, as we shall see, he thinks it essential for all religion that God be anthropomorphic at least in being benevolent.

Now the battle line between Philo and Cleanthes is drawn. Cleanthes tries to infer what he can about God while Philo deftly and Demea unwittingly keep cutting the ground from under him.

Part X is another stage, and a crucial one, in a series of Cleanthes' attempts to infer what he can about God. At this stage Cleanthes is trying to establish God's benevolence; God's existence has not been questioned, nor will it be. We must notice where Philo enters the discussion and where he aims:

And is it possible, Cleanthes, that after all these reflections, and infinitely more, which might be suggested, you can persevere in your anthropomorphism, and assert the moral attributes of the Deity, his justice, benevolence, mercy, and rectitude, to be of the same nature with these virtues in human creatures? His power, we allow, is infinite: Whatever he wills is executed: But neither man nor any other animal are happy: Therefore he does not will their happiness. His wisdom is infinite: He is never mistaken in choosing the means to any end: But the course of nature tends not to human or animal felicity: Therefore it is not established for that purpose. Through the course of human knowledge, there are

⁶ Huxley (*op. cit.*) says this makes us doubt whether Philo ought to be taken as Hume's mouthpiece because in the *Treatise of Human Nature*, Book I, Part III, Sections III and XIV, Hume affirms that "there is no absolute nor metaphysical necessity that every beginning of existence should be attended with such an object" [as a cause]; and again, that it is "easy for us to conceive any object to be non-existent this moment and existent the next, without conjoining to it the distinct idea of a cause or productive principle." But Hume explains his meaning in his letter to John Stewart: "I never asserted so absurd a Proposition as that any thing might arise without a Cause: I only maintain'd, that our Certainty of the Falshood [*sic*] of that Proposition proceeded neither from Intuition or Demonstration; but from another Source. That Caesar existed, that there is such an Island as Sicily; for these propositions, I affirm, we have no demonstrative nor intuitive Proof. Woud [*sic*] you infer that I deny their Truth, or even their Certainty? There are many different kinds of Certainty; and some of them as satisfactory to the Mind, tho perhaps not so regular, as the demonstrative kind." *The Letters of David Hume*, vol. I, ed. J. Y. T. Greig (Oxford, 1932), p. 187.

⁷ For a clear statement of this point and a helpful examination of the question "Does God exist?" see Paul Ziff, "About God" in *Religious Experience, and Truth*, ed. Sidney Hook (New York, 1961), pp. 195–202.

⁸ I agree generally with Professor Butler (*op. cit.*) when he says "Philo's entire criticism of the argument from design should be viewed as an attempt, not to deny that God exists, but to break down Cleanthes' initial opinion that theological beliefs may find rational support in the recognition of evidence" (p. 87); and when he says, "The *Dialogues* are an attempt to work out precisely how much or how little is involved in this concession" [that God exists] (p. 92). But I prefer my formulations of these two points as more specific for Part X.

not inferences more certain and infallible than these. (Hume, p. 66.)

Hume has been criticized for not seeing that God may have had to allow suffering for some reason or other; and, this being so, Philo's reasoning disproves neither God's existence, nor God's benevolence, nor God's omnipotence.⁹ But Philo asks this: "In what respect, then, do his benevolence and mercy resemble the benevolence and mercy of men?" (Hume, p. 66). There is no reason to suppose Hume thought Philo's reasoning disproved anything except that the course of nature was an adequate basis for saying the moral attributes of God are the same as those of humans.

Philo admits to Cleanthes the reasonableness of ascribing a purpose to nature, but he denies that the purpose is to benefit either man or beast:

You ascribe, Cleanthes, (and I believe justly) a purpose . . . to nature. But what . . . is the object of that curious artifice and machinery, which she has displayed in all animals? The preservation alone of individuals and propagation of the species? It seems enough for her purpose, if such a rank be barely upheld in the universe, without any care or concern for the members that compose it. (Hume, p. 66.)

Cleanthes sees the seriousness of Philo's attack, and he sees it for what it is—neither an assertion that God does not exist, nor an assertion that God is not benevolent, but an attack upon the idea that the course of nature is a basis for asserting God's benevolence.¹⁰ So Cleanthes tells Philo: "If you can . . . prove mankind to be unhappy or corrupted, there is an end at once of all religion. For to what purpose establish the natural attributes of the Deity, while the moral are still doubtful and uncertain?" (Hume, p. 67).

Now Philo is in a position not entirely satisfactory to a sceptic; he is asked to prove something and, scepticism aside, something difficult, if not impossible to prove. His next move is important, in fact, the crux of this dialogue, but before he can make it, Demea interrupts to state his theodicy. The interruption helps Philo strategically, and it is a chance for Hume to interject a representative line of thinking which he thinks must be cut down

before the full impact of Philo's scepticism can be appreciated.

Demea's theodicy is the so-called "porch view":

This world is but a moment in comparison of eternity. The present phenomena, therefore, are rectified in other regions, and in some future period of existence. And the eyes of men, being then opened to larger views of things, see the whole connection of general laws, and trace, with adoration, the benevolence and rectitude of the Deity. . . . (Hume, p. 67.)

This theodicy directs attention to the broad and eternal view of existence, and it is general enough to represent most theodicies. It makes either of two assumptions. One characterizes our view as being so limited that, even though we think we suffer, we really do not. Presumably, a broader or longer view of things would disclose our error. The other is that our view is so limited that, even though we really suffer when we think we do, we do not see that we must suffer for the sake of a greater good, either for ourselves or for the whole world. On the one hand, there really is no evil; on the other, evil is necessary.¹¹

Whichever assumption Demea makes, his theodicy will not withstand Cleanthes' blow against it: "Whence can any cause be known but from its effects? To establish one hypothesis upon another is building entirely in the air: and the utmost we can ever attain, by these conjectures and fictions, is to ascertain the bare possibility of our opinions; but never can we, upon such terms, establish its reality" (Hume, p. 68). So this theodicy and all others like it cannot interfere with Philo's line of argument. Demea has no naturalistic evidence to claim that there are regions and a period of future existence where the present evil phenomena are rectified.

And, while Cleanthes admits that these conjectures and fictions ascertain the bare possibility of our opinion, thus recognizing that Philo's trilemma does not logically exclude the possibility of a benevolent God, still these conjectures presuppose that God is a moral agent—that he will, because of his nature, rectify the present evils of man—when this is precisely the point at issue.

⁹ Nelson Pike, *op. cit.*

¹⁰ This strictly parallels the argument Hume puts forth in the *Inquiry Concerning Human Understanding*, sect. XI. See the explication of it by Antony Flew in *Hume's Philosophy of Belief* (New York, 1961), pp. 222–223.

¹¹ The distinction between these two assumptions is seldom made. Berkeley seems to think evil is an illusion, but he uses the second notion, that evil is necessary, to explain the illusion (*Principles of Human Knowledge*, Part I, sect. 153). Berkeley's view resembles Demea's in form and in lack of form. Leibniz, of course, uses the second, more common notion (*Theodicy: Summary of the Controversy Reduced to Formal Arguments*, Objection I, Answer).

Cleanthes knows Demea cannot argue that God will, in some unknown way and for some unknown reason, validate man's suffering; nor can he base his argument on God's benevolence when he does not know that God is benevolent—when, in fact, he is trying to prove that benevolence in the presence of suffering makes it doubtful. So he says to Demea, "The only method of supporting divine benevolence (and it is what I willingly embrace) is to deny absolutely the misery and wickedness of man" (Hume, p. 68).

And now Philo is ready for his crucial move. He says:

I . . . must admonish you, Cleanthes, that you have put the controversy upon a most dangerous issue, and are unawares introducing a total scepticism into the most essential articles of natural and revealed theology. What! no method of fixing a just foundation for religion unless we allow the happiness of human life, and maintain a continued existence even in this world, with all our present pains, infirmities, vexations, and follies, to be eligible and desirable! But this is contrary to everyone's feeling and experience; it is contrary to an authority so established as nothing can subvert. No decisive proofs can ever be produced against this authority; nor is it possible for you to compute, estimate, and compare all the pains and all the pleasures in the lives of all men and of all animals; and thus, by your resting the whole system of religion on a point which, from its very nature, must for ever be uncertain, you tacitly confess that that system is equally uncertain. (Hume, pp. 68-69.)

Philo has turned the trick without proving or disproving anything, not even that mankind is unhappy. Cleanthes, himself, has shown that the only way to establish the divine benevolence and, consequently, natural religion itself is to stand in a quagmire.

Philo then takes the argument to a second stage and even allows "what can never possibly be proved"—that human happiness exceeds its misery. This takes Cleanthes nowhere because from infinite power, wisdom, and goodness we should reasonably expect no misery in the world at all. The only escape from logic so solid and decisive is to deny that we know anything about these matters. This, says Philo, he has maintained from the beginning of the discussion (Hume, p. 69).

Philo next takes the argument to a third stage and even grants that pain or misery in man is compatible with infinite power and goodness, even in the ordinary sense of these attributes. Even this takes Cleanthes nowhere, for he must prove that the Deity has these attributes from the present mixed and confused phenomena, and from these alone. Even if there were no evil, Cleanthes would confront sufficient difficulties because the phenomena are finite. Actually, there is evil and the phenomena are mixed (Hume, p. 69). Clearly, Philo is not arguing against God's existence, for then allowing the compatibility of evil and the divine attributes would amount to capitulation.

In Part XI Cleanthes tries to avoid Philo's conclusion and preserve the human analogy: "Supposing the author of nature to be finitely perfect, though far exceeding mankind, a satisfactory account may then be given of natural and moral evil . . . benevolence, regulated by wisdom and limited by necessity, may produce just such a world as the present" (Hume, p. 71). But this supposition allows merely the compatibility of evil and divine benevolence. Philo has already in effect shown it of no avail in Part X, where, for the sake of argument, he allows the compatibility of evil with *infinite* power and goodness. In Part XI he has merely to say, "Conjectures, especially where infinity is excluded from Divine attributes, may perhaps be sufficient to prove a consistency, but can never be foundations for any inference" (Hume, p. 73). And inference has been the matter all along.

So with God's moral character at stake, with Cleanthes' attachment of the fate of natural religion to the fate of God's benevolence, with Cleanthes' demolishing blow against Demea's theodicy and all others like it ("to establish one hypothesis upon another is building entirely in the air"), and with the necessary foundation of natural religion shown to be in principle unprovable, even allowing the compatibility of evil and divine benevolence, Philo justly considers his case logically tight. While this is not a demonstration that God does not exist, nor that God is not benevolent, it is a perfect triumph for the sceptic on perhaps the crucial issue for natural religion.

AMERICAN PHILOSOPHICAL QUARTERLY

Edited by
NICHOLAS RESCHER

With the advice and assistance of the Board of Editorial Consultants:

William Alston
Alan R. Anderson
Kurt Baier
Lewis W. Beck
Richard B. Brandt
Roderick M. Chisholm
L. Jonathan Cohen
James Collins
Michael Dummett

James M. Edie
Peter Thomas Geach
Adolf Grünbaum
Carl G. Hempel
Jaakko Hintikka
Raymond Klibansky
Benson Mates
John A. Passmore
Günther Patzig

Richard H. Popkin
Wesley C. Salmon
George A. Schrader
Wilfrid Sellars
J. J. C. Smart
Wolfgang Stegmüller
Manley H. Thompson, Jr.
G. H. von Wright
John W. Yolton

VOLUME 3/NUMBER 2

APRIL 1966

CONTENTS

- | | | | |
|--|-----|--|-----|
| I. RICHARD ROUTLEY AND VALERIE
MACRAE: <i>On the Identity of
Sensations and Physiological Occur-
rences</i> | 87 | IV. JOEL FEINBERG: <i>Duties, Rights, and
Claims</i> | 137 |
| II. DIOGENES ALLEN: <i>Motives, Ration-
ales, and Religious Beliefs</i> | 111 | V. RICHARD M. GALE: <i>McTaggart's
Analysis of Time</i> | 145 |
| III. FREDERICK A. SIEGLER: <i>Lying</i> | 128 | VI. A. N. PRIOR: <i>Postulates for Tense-
Logic</i> | 153 |
| | | VII. JOSEPH OWENS: <i>The Grounds of
Universality in Aristotle</i> | 162 |
-

PUBLISHED BY BASIL BLACKWELL WITH THE COOPERATION OF THE UNIVERSITY OF PITTSBURGH

AMERICAN PHILOSOPHICAL QUARTERLY

POLICY

The *American Philosophical Quarterly* welcomes articles by philosophers of any country on any aspect of philosophy, substantive or historical. However, only self-sufficient articles will be published, and not news items, book reviews, critical notices, or "discussion notes."

MANUSCRIPTS

Contributions may be as short as 2,000 words or as long as 25,000. All manuscripts should be type-written with wide margins, and at least double spacing between the lines. Footnotes should be used sparingly and should be numbered consecutively. They should also be typed with wide margins and double spacing. The original copy, not a carbon, should be submitted; authors should always retain at least one copy of their articles.

COMMUNICATIONS

Articles for publication, and all other editorial communications and enquiries, should be addressed to: The Editor, *American Philosophical Quarterly*, Department of Philosophy, University of Pittsburgh, Pittsburgh, Pennsylvania 15213.

OFFPRINTS

Authors will receive gratis two copies of the issue containing their contribution. Offprints can be purchased at low cost through arrangements made when checking proof.

SUBSCRIPTIONS

The price *per annum* is six dollars for individual subscribers and ten dollars for institutions. Checks and money orders should be made payable to the *American Philosophical Quarterly*. Back issues are sold at the rate of two dollars to individuals, and three dollars to institutions. All correspondence regarding subscriptions and back orders should be addressed directly to the publisher (Basil Blackwell, Broad Street, Oxford, England).



I. ON THE IDENTITY OF SENSATIONS AND PHYSIOLOGICAL OCCURRENCES

RICHARD ROUTLEY AND VALERIE MACRAE

I

THE main hypothesis to be considered is, in initial formulation:

(A) *Sensations are physiological occurrences*

"Physiological" is used in its normal sense, so that physiological occurrences include both bodily occurrences like the damaging of tissues and neurophysiological occurrences. The frequently discussed formulation "Sensations are brain processes" over-restricts the class of physiological processes relevant to sensations.¹ When (A) is asserted it is generally intended as an identity. To make the identity explicit (A) needs reformulation; for at face value it expresses an *inclusion*. Identity is a symmetrical relation; but the truth-value of "Physiological occurrences are sensations" may well differ from that of (A). Other indeterminacies in (A) appear as soon as we start asking about equivalent, or synonymous, reformulations. Improved formulations which avoid some of these problems are:

(B₁) *Sensations of sort $\hat{h}(g)$ that belong to or are present in x are identical with physiological occurrences of sort g in x , for any x and any g .*

(C₁) *x has a sensation of sort $h(g)$ iff a physiological occurrence of sort g takes place in x , for any x and any g .*

What, however, are the ranges of the variables? Consider the variable ' x '; an obvious choice is to let this variable serve as placeholder for referring expressions which refer to particular animals, including humans. The predicate-variable ' g ' serves as placeholder for certain descriptions (most details of which are not known) of physiological processes; ' h ' stands for a function which correlates described physiological occurrences with sensa-

tions; and so ' $h(g)$ ' ranges over descriptions of sensations. Actually, the relation between a subclass of physiological occurrences and sensations may not be so simple that it can be expressed by a single-valued function, or even by an explicit function like h . (C₁) has the form of a standard type of reduction sentence; whether it is or not depends on the tricky issue as to whether it contains the requisite observational and theoretical expressions (see Sect. II).

Before discussing the relevant senses of "iff" and "are identical with," it is important to clarify the sense and designation of "sensations" and to distinguish sorts of sensations. A quick glance at the etymology of "sensation" (*vide* the OED) shows that the word has been used in quite a wide way since about 1500 to cover items like feelings, e.g., of pride and elation, and suppositious items like visual, tactual, kinaesthetic sense-impressions or -data, as well as physically localizable, often disturbing, person-limited bodily sensations. To avoid confusion and the necessity of making qualifications "sensation" will be used, from now on, in the sense in which it is restricted to refer to (occurrences of) items of this last cluster, that is, to sensations proper. Included in sensations proper are: bodily pains of all sorts; itches, ticklings, shivers, numbness, and blackout sensations; internal pressure, orgasm, giddiness, thirst, nausea, seasickness, suffocation, dazzle, and after-image sensations. The distinguishing marks of sensations proper explained below are vital because:

(i) These marks serve to delimit the cluster and reveal important similarities between various different sorts of sensations, e.g., between sensations as different as bodily pains, giddiness, and the having of after-images.

(ii) For each group of features there are many mental items, such as emotions, thoughts, volitions, and mental dispositions, which differ from

¹ Similarly, central state materialisms soon have to be widened to include other bodily processes and behavior, e.g., that of endocrine systems which seem important in connection with emotions. Once the extension is made some of the barriers separating materialisms and analytical behaviorism are down.

sensations in respect to these features. Some of the features also serve to isolate sensations from perceptual observations.

(iii) These features make identity hypotheses look much more plausible and more testable when restricted to sensations than when extended to cover all mental items. The restricted hypotheses are more amenable to testing, and (as it already seems) we can establish an accurate temporal correspondence between and also similarities of patterns of sensations and physiological occurrences.

Restricted identity-hypotheses are more plausible for similar reasons (developed in setting out the marks of sensations). This is so for three reasons: (1) because of the precision with which the duration of sensations can be marked out; (2) because of their bodily locatability; and (3) because of the many ways in which they can be assimilated to physiological occurrences. For these reasons, restricted hypotheses like instantiations of (B_1), are not to be confused with full materialist hypotheses or accounts of mind, which endeavor to make a reduction—of a fairly uniform character and satisfying some or other (though often it is not clear which) of a range of possible reduction relations—of *all* mental items to physical or material data of *some* sort. Restricted hypotheses are *compatible* with more sophisticated many-strand theories different from materialism, indeed with analytical behaviorism which is not committed to offering a behavioristic account of sensations; though vindication of (A) would make serious inroads into dualisms.

So there are two suppositions we put up:

(a) Materialisms, especially central-state theories, are more plausible for sensations, just because of some special features of sensations, than for most (other) mental items; if it fails here then it fails entirely. Moreover, if it runs into difficulties over sensations then it is much more likely to run into difficulties, some of them transferable from the sensation case, over emotions, thoughts, and decisions.

(b) Sensations proper are *not* correctly classified as *mental* occurrences. Admittedly it is fairly standard nowadays to classify sensations, though not reflex actions, as mental occurrences; but the classification does not have a good pedigree (cf. Greek classifications and Cartesian difficulties over animals), and when in the past sensations did get in they tended to have second-class status. It is possible, by taking into account different senses of

"mind" and of "mental" and features of sensations, to develop something of a case for their reclassification. For instance, sensations don't have much more relevance to a person's mental, intellectual, or emotional states than might some of his physical characteristics, e.g., having lost a limb. A person who had no sensations could still have a mind. Also the question "Do rats have minds?" is a conflict-question but the question "Do rats have sensations?" is not. Consider, too, details introduced in discussing someone's mental qualities, whether he has a good mind, what his mental state is. The conventional way of settling this conflict issue over whether sensations are mental occurrences or not, seems to draw much of its strength from dubious stream-of-consciousness or sensationist views on what it is to have a mind. If (b) is defensible, sensations, and relatively incorrigible occurrences, are very far from being paradigms of mental occurrences; and establishing of (B_1) would go little distance toward establishing a full materialist account of *mind*, though, if (a) is correct, refuting (B_1) would go a good way toward refuting it.

Important distinguishing marks of sensations, on which fuller defense of the restricted identity claim relies, are then:

(1) Sensations are items that we *have* at given times or over periods of time. They last fairly definitely specifiable periods of time. In general, one can say, with a high degree of precision, of a sensation that it's just started or that it's now gone. Sensations are *occurrences* or *processes*. They are not events in the narrow sense of "event," neither are they like achievements which have a time of happening but no duration. Sensations are occurrences or perhaps, if they are sufficiently frequent or dominant, "states." They are episodic and not dispositional though a person may be disposed to or liable to get certain (sorts of) sensations. Because sensations are occurrences, it is the *having* of after-images and not after-images themselves, i.e., processes and not the products, that count as sensations.

(2) All sensations either have a *bodily location*, which may, as in tingling or fever sensations, be quite extensive, or else are intimately connected, like after-imaging, with a bodily condition. To guarantee this point in the case of pains we take over the traditional distinction between mental pains, like anguish, which are better classed with emotions, and bodily pains. This distinction can be sharpened by the adoption of physiological criteria for real bodily pains.

(3) They happen to us; they are not something we do. In this respect especially, sensations may be likened to reflex actions, with which they are often connected, as for instance withdrawal from a cause of pain and avoidance of a source of dazzle. True, with all of them there are recipes or procedures for producing them: but having carried out these procedures and set up appropriate conditions, there is no stopping having them, except that in some cases there are counter-procedures we can use to prevent them provided the procedures are initiated early enough. We can have sensations accidentally or without intending to and when we don't want to have them; and they aren't produced just by *deciding* to have them without setting up the appropriate physical conditions connected with their occurrence. Consequently they are not put down to our agency, we are not held *responsible* for having them, though we may be held responsible either for causing or setting up the conditions for them either in ourselves or, more often, in others. Thus even the having of ordinary sensations is unconnected (except in contingent ways, e.g., in psychosomatic disorders) with personality and intellectual traits.

(4) Although some people are better than others at detecting, locating, and describing their own, and even others' sensations and sometimes at producing them, the having of sensations is not at present an exercise of skill or care or a matter for training; nor in general are their occurrences a part of or the result of discrimination, detection, or the gaining of information. Thus there is little scope for mistake, little liability to correction with regard to the having of sensations. These features set the having of sensations apart from perceptual observation, from seeing, watching, tasting, etc., where training is relevant, where skill can be gained and may be needed, where mistakes are common enough, some illusions are normal, and achievements occur. Through visual, tactual, and other processes we often gain information about external items and their relations, and sometimes about internal items; but the having of sensations does not provide us directly with such information though, since certain sensation patterns are regularly produced under certain conditions; we can sometimes get data about external conditions in this way. Despite these differences the precise boundary between sensations and sense perception may remain problematic if it is not noticed that kinaesthetic receptors serve a dual purpose (thus heat receptors appear both in physiological

accounts of shivering and fever sensations and in the causal-physiological story of the way we detect the temperatures of external items). It should also be noticed that much of our vocabulary for describing sensations is transferred from that used to specify observable processes or features (consider descriptions of pains and after-imaging). With sensations we do not do better than achieve comparisons with external items, comparisons which already require perceptual observation of such items. Sensations do not in fact have external correlates; no after-images on walls, no pains in red-hot elements.

(5) First-person sensation reports are *relatively* incorrigible; havers are not likely to be corrected. Mostly sensations are readily identifiable to the haver. Very many of them are describable in detail as to their sort, pattern, history, changes, intensity, duration, and location or apparent location. Indeed, a person who has sensations cannot help noticing that he has them if he has them with any intensity and provided there are no sufficiently important distractions. The qualifications exclude rare exceptions such as the wounded soldier or the person who has undergone lobotomy.

(6) When the appropriate conditions—conditions which vary with the sort of sensation but are standard and specifiable for most sorts of sensations—are fulfilled everyone who is in a normal physiological condition in the relevant respects has the sensation. After-imaging and angina pains provide typical examples.

(7) If asked to explain the occurrence of specific sensations we usually resort to *physiological-type-explanation* and indicate some physical source and cause. It is worthwhile separating out several explanations of this type:

(a) Explanations referring to extra-bodily agents or objects which explain the *onset* of sensations but not, as a rule, their continuation; e.g., Jack hit you with a shovel, and now you are seeing spots of light. (In certain cases, with rapidly recovering tissues, etc., the continuation of a sensation may be explained by continual contact with the external source.)

(b) Explanations referring to bodily non-cerebral regions and processes in or features of these regions, as when the tissues in your shoulder are damaged (through the impact of the shovel), cones of your eyes are bleached (through staring at the light). When combined with laws of such forms as, "tissues damaged (fatigued, . . .) in respect *R* take under conditions *C* time *t* to recover, exhibit pat-

tern P in recovering," facts under (b) can be used to explain the continuation of sensations, their duration, their approximate intensity, and their pattern. Consider after-imaging: given details of the structure of the eye, in particular the photochemistry of cones and rods and of the optical nerve; given recognized physical and chemical generalizations, and given in each case relevant initial conditions like properties of the item observed, its relation to the observer, and period of observation, one can give a fair explanation of, and can predict, after-imaging and features of it. One can predict the duration and appearance, size, shape, brightness, and saturation of the images had. To complete our explanations we turn to:

(c) Explanations referring to neurophysiological occurrences which tell a story, generally starting with reactions of nerve fibers in regions referred to in (b) and ending with processes occurring in the brain, and in particular in the relevant sensation areas, e.g., pain-centers or c-fibers. In the case of psychosomatic sensations (a) and (b) may not be relevant; though diagnosis of stresses on the patient may be.

A complete context-independent explanation of specific sorts of sensations, which should in the end explain most of their features, combines (a), (b), and (c). There seems no upper limit short of explanation of *all correctly* observed and reported features of sensations as our explanations progressively improve with increased physiological data and assimilation of it under laws. These explanations would at least discern regular correlations between given sorts of sensations and physiological occurrences and result in control over sensations. By applying the explanations sensations could be induced, prolonged, shortened, or alleviated. But when observations of relations between correlated processes (strictly observations of relation-instances) are precluded or seriously limited (as here), regular correlations, as linked with explanations and connected under time-directed laws, together with related manipulative control, constitute first-rate grounds for the presence of causal relations. Therefore, given all these, the physiological explanations would be what they seem to be—causal explanations. It is extremely difficult, perhaps impossible, to obtain general explanations or characterizations of the causes of sensations without introducing

physiological matters. Asked to explain³ what caused a particular after-image we may say "Looking at the one light over there too long"; but asked to explain after-imaging in general it is not usually satisfactory to say "After-imaging is caused by looking at bright objects too long." Physiological explanation is needed.

II

Consider hypotheses (B₁) and (C₁). Abbreviating "sensations of sort $h(g_0)$ in x_0 " by "SENGHS" and "physiological occurrences of sort g_0 in x_0 " by "PROGS," we may write an instantiation of (B₁):

- (B) SENGHS are identical with PROGS; or
 $\hat{x}: (\text{SENGH}(x)) = \hat{x}: (\text{PROG}(x))$
- (B) states a contingent class identity rather like
- (D) Humans are featherless plantigrade bipeds; or
- (D₁) Genes are (large parts of) DNA molecules; and
not like
- (D₂) Brothers are male siblings

where (D₂) yields an analytic statement. Unless the identity in (B) is formulated more explicitly problems arise as to what statement (B) is intended to yield. How can (B) be made more explicit? A recent move is to invoke a sense/reference distinction; to distinguish

(E₁) "SENGHS" has the same sense as "PROGS,"
 from

(E₂) "SENGHS" has the same referents as "PROGS"

and to equate (B) with (E₂). (E₂), despite its merits as a reformulation, has some drawbacks: first, it isn't obvious how extensive the referent-class is; second, "same" also appears in (E₂). In fact the intended referent-class, e.g., of "SENGHS" is the class of all referents which exist at some time, i.e., the total extension; and this last class is *independent* of the sense of "SENGHS" except that the class is a subclass of the possible extension which is logically determined by the sense of "SENGHS." Consequently we can *consistently adjoin* to a specification of the total extension of a referring expression (e.g., the class of all sometime actual featherless bipeds or "humans") a specification of the sense of the

³ Here "explanation" is being used in a common way, so that explanations are context-dependent, and what constitutes a satisfactory explanation, like a satisfactory description, is relative to the purposes, interests, and background knowledge of speaker and audience.

expression (e.g., "animals belonging to *homo sapiens*" of "humans"). In the same way *certain* specifications of the sense of "(bodily) pain"—e.g., the quite unsatisfactory (E_1) (with a suitable choice of g) or the more satisfactory

(E_3) "(Bodily) pain" has the same sense as "sensation of a certain sort, viz., the distressing sort typically felt when physiologically hurt,"

can be consistently adjoined to a specification like (E_2) (with the same choice of g) of the total extension of "pains." But though we can consistently couple specifications of sense like (E_3) with (E_2) we don't have to. Therefore a contingent identity theory with a main hypothesis like (B) need not provide an analysis, topic-neutral or otherwise, of sensations or specifications of the senses of "bodily pain," "after-imagining," and so on.

(E_2) can be more explicitly formulated as

(E_4) The total extension of "SENSEHS" is identical with the total extension of "PROGS."

The only material advantage of (E_4) over (B) is that it is clear that the sentence yields a contingent statement; and this could be made clear in (B) by replacing "identical" by "contingently identical" (call the result (B_2)). Unfortunately (E_4) leaves us with the same major problems as (B) or (B_2): for criteria for and conditions on identity are crucial. These criteria a sense/reference distinction does not automatically supply. Although it is certainly valuable to have a sense/designation distinction, or at least to have the tie-up between contingent identity and sense clarified, in order to reinforce the argument at various points, the subsequent discussion could be accomplished without appeal to such distinctions. Sense/reference distinctions can be by-passed.

The usual (Leibnitz') definition of *identity*,

- (1) $x = y$ FOR $(\forall f) \cdot f(x) \equiv f(y)$,
 where ' f '³ is a sentential function (or predicate),
 leads at once to
 (2) $[x = y] \supset (\forall f) [f(x) \equiv f(y)]$,

the principle of indiscernibility of identicals; and to
 (3) $(\forall f) [f(x) \equiv f(y)] \supset [x = y]$,

the principle of identity of indiscernibles. The argument-range of ' x ' includes class- as well as individual-expressions. For almost all purposes, outside limited contexts like extensional logics where the class of predicates is restricted, (1) is much too strong. In particular, if adopted as a criterion of identity it would eliminate all scientific identifications and even many identifications within mathematics. For consider a tentative identity: $x_1 = y_1$, e.g., an identity of the sort made in scientific reductions like

(F_1) The temperature of an (ideal) gas is the mean kinetic energy of its molecules (strictly: $T = \frac{2}{3k}E$);

and consider the property specified by "is known by Arthur to be identical with x_1 " or abbreviated " $K_a(\dots = x_1)$," as in the statement "is known by Arthur to be identical with the temperature of a gas." If Arthur does not know much science we have as true statements $[K_a(x_1 = x_1)]$ and $[\sim K_a(y_1 = x_1)]$, whence by (1), $[x_1 \neq y_1]$ is true. There is an obvious way out of this impasse, as well as out of some standard difficulties raised by (3) over items with identical properties apart from spatio-temporal or temporal location in universes with limited accessibility, and by (2) once modal principles are introduced into predicate logics or set theories. First, limit the class of properties admitted and second, restrict conditions under which the definitions or implications hold. To achieve the first of these moves some preliminary terminology is valuable.

A property f is *intensional* or *non-extensional* iff either (i) ' f ' has a significance-range which includes the statement-expressions or infinitival expressions represented by ' p ' and ' q ' but f fails to satisfy the condition

(Ext) $(\forall p, q) \cdot p \equiv q \supset f(p) \equiv f(q)$

or (ii) ' f ' does not have such a significance-range but ' f ' contains a subpredicate ' g ' and g satisfies

³ Angle quotes represent the quotation function 'qu', which is so defined that for any given argument-value the value of the function is the quotation-expression of that argument. A statement-expression is an expression of the form "that..." where the blanks are filled by statemental sentences. Statemental sentences can also belong, derivatively, to the significance-range of ' f '. An infinitival-expression is an expression of the form "to..." where "to" is part of an infinitive, e.g., "to go home," "to be Pope." Some such expressions are needed if properties like "Tom wants..." "Tom liked..." are to be classed as intensional under the definition and grammatical ugliness such as "Tom wants that he is Pope" avoided. The definitions are phrased negatively in order that such properties as "... is tall," "... is blue," "lasts t seconds" may qualify as intensional, strict, etc. These properties qualify because their predicates fail to satisfy a significance condition. On approximately this point see S. Halldén, *The Logic of Nonsense* (Uppsala, 1949), pp. 47-48.

(i); otherwise f is *extensional*.^{3a} We abbreviate this definition: A property f is *intensional* if f fails on (Ext); otherwise f is *extensional* (abbreviated: *ext*). The definitions which follow are presented in a similar abbreviated form. A property f is *non-strict* if f fails on

(Str) $(\forall p, q) . \Box(p \equiv q) \supset . f(p) \equiv f(q)$;
and *strict* otherwise. A property f is *non-regular* if f fails on

(Reg) $(\forall p, q) . (\forall x) Kx(p \equiv q) \supset . f(p) \equiv f(q)$,
where ' Kx ' symbolizes " x knows (that)"; and *regular* (*reg*) otherwise. Thus " a knows . . .," " a believes . . .," and extensional properties like " a is taller than a " are regular, but " a hopes" is non-regular. "It is necessary . . .," "It is possible . . .," and all extensional properties are strict, but "Tom believes . . .," "It is possible Tom wants . . ." are non-strict. A property f is *modal* iff f is strict but non-extensional; *narrowly propositional* iff f is regular but non-strict. Intensional properties include not only "... is known by a ," " a believes . . .," "is about a ," but—most important—"... is about something private," " a noticed . . .," "... belong to the surroundings of" and " a 's statement . . . is incorrigible." The non-extensionality of the last property can be shown by applying the definition to the equivalence:

[I have a pain] \equiv [The man born in x_3 who has done f_3 has a pain].

Consider now the following identity definitions:

(4) $x = y$ FOR $(\forall \text{ ext } f) . f(x) \equiv f(y)$,

i.e., in classical logics:

$(\forall f) . \text{ext}(f) \supset . f(x) \equiv f(y)$.

^{3a} Clause (ii) of the definition of *intensional* is designed so as to include as intensional properties like "Arthur knows that Bill is the same individual as . . .," " a is frightened of . . ." For the relevant predicates contain non-extensional subpredicates; in the examples "Arthur knows" and " a is frightened" respectively. Strictly the definitions of *intensional*, *non-strict*, etc., are defective in the form given. To repair the defect 3-valued significance logic has to be used explicitly. The definition of *intensional* is repaired by replacing (Ext) by

(i) $(\forall p, q) . \Sigma^<f(p)> \& \Sigma^<f(q)> \supset . p \equiv q \supset . f(p) \equiv f(q)$.

To be exact similar amendments should be made to (Str) and (Reg). In these modified conditions ' Σ ' symbolizes "is significant" (Halldén's '+'), and ' \equiv ' represents a 3-valued equivalence which is true if its components have the same value and false otherwise. Alternatively (i) can be replaced by

(ii) $(\forall p, q) . p \equiv q \supset . \Pi(f(p) \equiv f(q))$,

where $[\Pi p]$ is false when $[p]$ is false and true otherwise.

⁴ For, using Feys's system T (a system displayed in A. N. Prior, *Formal Logic*, 2nd edition [Oxford, 1962], p. 312) and classical predicate logic, it follows:

1. $[x = x]$
2. $\Box[x = x]$
3. $[x \equiv y] \supset [f(x) \equiv f(y)]$
4. $[f(x)] \supset [[x \equiv y] \supset f(y)]$
5. $\Box[x = x] \supset [[x \equiv y] \supset \Box[x = y]]$
6. $[x \equiv y] \supset \Box[x = y]$
7. $[\sim \Box[x = y]] \supset [x \neq y]$
8. $\Delta[p] \equiv \Diamond_{\text{Dt}}[p] \& \Diamond[\sim p]$
9. $[\Delta[x = y]] \supset [\sim \Box[x = y]]$
10. $[\Delta[x = y]] \supset [x \neq y]$.

from (4) or any other plausible definition of ' $=$ '.

from 1, by the rule of T .

provided f is strict; from (5).

from 3, by propositional logic.

substituting ' $\Box[x = . . .]$ ' which is a strict predicate for ' f ' in 4.

from 2, 5.

from 6.

(definition of " p " is contingent).

from 8.

from 7, 9.

(5) $x \equiv y$ FOR $(\forall \text{ strict } f) . f(x) \equiv f(y)$.

Other identity criteria may be obtained by considering different classes of properties. All these identity criteria may be connected by the scheme for R-identity.

(6) $x =_R y$ FOR $(\forall f \in R) . f(x) \equiv f(y)$,

where R is a class of properties which includes at least all extensional properties. For example *extensional identity* defined by (4) is obtained by taking R to coincide with the class of extensional properties; *strict identity* defined by (5) by taking R to be the class of strict properties; *Leibniz' identity* defined by (1) by taking R to be the class of all properties.

Even strict identity is too strong if (B) or *any contingent identifications* are to survive. Consequently any identity criterion such as Leibniz' identity which entails strict identity is also too strong for contingent identities like (B). For if it is only a contingent matter that x is identical with y then x is not strictly identical with y .⁴ Since many sentences containing "identical with," in particular (D), (F₁), and (B), only yield synthetic statements strict identity is too strict. We must relax our requirements below those of Leibniz', and step down at least to a class R of properties from which most intensional properties are excluded. The example of Arthur shows that " $Ka . . .$ " must be excluded from R . Similar examples can be constructed to show not merely that all narrowly propositional properties but that very many other intensional (mental) attributes have to be excluded from R . Further examples can be designed

to show that many impure properties (those whose specifying predicates include essentially quotation) and [as in step 5, footnote 4] many modal properties which can be significantly attributed to individuals fail to qualify for *R*. Because only extensional properties clearly qualify for *R*, because of the indeterminacy of the class of further properties which might qualify for *R*, and because (4) is sufficient to retain important substitution (*salva veritate*) principles and sufficient to guarantee identity of reference, we propose to step down to (4). (4) certainly seems to be very like what is required for a satisfactory contingent identification; any criterion much stronger would bring down most scientific identifications.

If Leibniz' or even strict identity were taken as a criterion of contingent identity (symbolized ' \cong ') it could be shown

$$x \cong y \equiv \Box(x \cong y)$$

and using Lewis' system S5 we obtain

$$\sim(x \cong y) \equiv \Box \sim(x \cong y).$$

Thus a contingent identity would be logically impossible. All true identity and difference statements would be analytic. Such a result is but a special case of the principle that all true statements are analytic, from which Leibniz claimed to deduce (1). Identity criteria (1) and (5) can be retained only if either all identities are analytic or else all predicates demarcate extensional properties. Both alternatives are false. Therefore (1) and (5) cannot be retained as contingent identity criteria. An identity criterion syntactically identical with the Leibniz criterion can, however, be obtained by pruning down the class of properties. Instead of redefining the word "property" let us call the properties which are so admitted *traits*. Many predicates which demarcate properties will not demarcate traits. So an important question is: Which properties are traits? If traits are identified with extensional properties, then (1) holds provided the *interpretation* is changed so that '*f*' now takes as values not properties but traits. Such a procedure, to keep up old appearances, provides only a thinly-disguised terminological variant of the course we have followed.

There are other ways of *apparently* avoiding replacement of (1) by (4), *extensionalizing moves*, which propose replacement of all intensional properties either (i) by (distantly related) extensional properties or (ii) by extensional and impure properties. There are also connected extensionalizing

techniques, used to retain (2), which come down to allowing substitutions only in extensional sentence contexts. But serious objections to both (i) and (ii) seem unlikely to be surmounted. Moreover (i) immediately concedes (4) for *all* identities and thus it strikes difficulties which adoption of (4) just for extensional identity avoids. For example, the important distinction between contingent and strict identity becomes difficult to draw, and escape routes from the Black dilemma [see below] are blocked. In case (ii) Leibniz-identity is qualified in the usual manner so as to exclude sentential functions which contain quotation essentially. If (ii) were viable this would include all intensional properties not already replaced by extensional analogues. Therefore (4) is conceded for all identities. Consequently (ii) encounters other difficulties, one of them being the problem of how to define non-extensional identities.

Direct arguments for criterion (4) run like this: If *x* and *y* are extensionally identical then they share all extensional properties. What needs to be shown, then, is that non-extensional properties are irrelevant to extensional identities. If *x* and *y* are extensionally identical it is only the extensions of '*x*' and '*y*' which are relevant in the identification. But non-extensional properties, such as modal properties, concern the sense of '*x*' and '*y*'. Therefore extensional identities are independent (in the sense explained earlier) of non-extensional properties; and (4) is established as an appropriate criterion for extensional identities. Lastly (4) can be made plausible by examples: consider sentences (G1)–(G5) below.

Given the adoption of (4), it follows that 'iff' in (C)-sentences only symbolizes a bi-conditional of approximately the same strength as a truth-functional bi-conditional and of much less strength than a mutual entailment or strict equivalence; and, given also classical two-valued logics, 'iff' does represent truth-functional equivalence. An instantiation of (C₂), namely

$$(C) \quad [\text{BENGH}(x_0)] \equiv [\text{PROG}(x_0)],$$

follows at once from the set-theoretic formulation of (B).

Adopting (4), i.e., opting for *extensional identity* at once eliminates several standard objections to (B); e.g., that various people can know about or understand details of one process but not the other, that the hypothesis is not good enough because, since the identity is only a contingent

matter, it could be false, that SENGHS are private but PROGS are not, that SENGHS are necessarily mental but PROGS are not, that SENGHS are sometimes noticed when PROGS are not, and that SENGHS require surrounding intensional practices and assumptions but PROGS do not. Adopting (4) also eliminates the Black dilemma⁵ constructed on the assumption that SENGHS are identical with PROGS: either there is no way of distinguishing SENGHS and PROGS because they are identical or there remain some irreducible mental properties. First, one horn of the dilemma breaks down once (4) replaces (1). SENGHS and PROGS can be distinguished even though identical using intensional properties. Since not all intensional properties are mental properties, mental properties are not indispensable in separating SENGHS from PROGS. Second, there is nothing in this separation to stop *extensional* identifications, e.g., of beliefs with physiological states. (B) is neutral with respect to contingent identifications, and even with respect to certain analyses of narrowly propositional properties in physicalistic, behavioristic, or other styles.

There are further difficulties which the adoption of (4) does not forestall; difficulties which face any *heterogeneous identification* or *reduction*. There are essential differences between

(I) *homogeneous identifications*; e.g., (D) or:

- (G₁) Scott is (identical with) the author of Waverley.
- (G₂) The Morning Star is (identical with) the Evening Star.
- (G₃) The successful general is the same person as the small boy who stole the apples.
- (G₄) $7 + 5 = 12$.

To be sure there are *important* differences between (D)s and (G)s on the one hand, and individual members of each of these collections on the other—and also between these and

(II) *heterogeneous identifications*; e.g., (F₁) or:

- (F₁) Lightning is an electrical discharge from ionized clouds of water-vapor to the earth.
- (F₂) Material objects are certain structured collections of atomic particles.
- (F₃¹) Rubber (caoutchouc) is atoms of carbon and hydrogen structured in polymers of isoprene.

- (F₄) Visible light is electromagnetic radiation of wavelengths between ca. 4000 and 7200 Å.
- (F₅) Pleasure is a certain sort of tension reduction.
- (F₆) Real numbers are classes of rational numbers.
- (F₆¹) $\sqrt{3} = \{x: \text{rat}(x) \ \& \ (x^2 < 3)\}$
- (F₇) Cardinal numbers are classes of classes.
- (F₇¹) 3 is the class of all classes which correspond 1-1 with the triple (x, y, z) .

The defining difference between (I) and (II) is sometimes explained in this way: the levels or types of the referring expressions on each side of any selected identity sentence in (I) are the same, while in (II) they are different. But it can be explained more satisfactorily thus: in (I) the same predicates and sorts of predicates can be significantly predicated of the referring expressions on each side of any selected identity-sentence, but in (II) *some but not all* predicates can be significantly predicated of both sides of an identity sentence.

As an illustration consider

- (F₈) the only object on the top of Everest = the flagpole erected by Hillary.

Since the requisite object might have been Hillary, and since significance is context-independent, it is significant to predicate of the left-hand side (LHS) of (F₈) but *not* of the right-hand side (RHS) of (F₈) such predicates as "was born in New Zealand," "likes ice cream," "is thinking about Vienna." Any number of examples like (F₈) can be concocted by taking descriptions of different types; as examples consider "The only thing ever to have scared Smith is the Benfield apparition," "The white dot on the horizon is Austin's house" and (F₉): the intersection of the classes of points described by the linear equations $(x = y)$ and $(x + y = 2) =$ the point with coordinates $(1, 1)$.

If points are individuals, it is significant to predicate of the LHS but not of the RHS of (F₉) "has one member" and "contains one point." Examples designed to show that (F₁)–(F₇) are properly classed as heterogeneous are, like most significance claims, controversial. We suggest: "Is the mean of the same property of its microcomponents" is significantly predicable of the RHS but not of the LHS of (F₁), and "is felt by most humans" and

⁵ This dilemma comes from an objection raised by Max Black to Smart's identity theory: see J. J. C. Smart, "Sensations and Brain Processes," *The Philosophical Review*, vol. 68 (1959). The objection is set out more explicitly in J. Shaffer, "Mental Events and the Brain," *The Journal of Philosophy*, vol. 60 (1963), pp. 160–166. For standard objections to identity hypotheses which adoption of extensional identity, or of contingent identity, knocks out, see N. Malcolm, "Scientific Materialism and the Identity Theory," *The Journal of Philosophy*, *ibid.*, pp. 662–663; Smart, *op. cit.*; Shaffer, *op. cit.*; J. Shaffer, "Recent Work on the Mind-Body Problem," *American Philosophical Quarterly*, vol. 2 (1965), pp. 81–105, and J. J. C. Smart, *Philosophy and Scientific Realism* (London, Routledge & Kegan Paul, 1963).

"occasionally fogs up the windows" of the LHS but not of the RHS of (F_1) . "Struck Tom," "is roughly chain shaped," "is regularly accompanied by thunder," and perhaps "is caused by an electrical discharge" and "is yellowish-blue," are significantly predicable of the LHS but not of the RHS of (F_2) , and "is sometimes elliptically polarized" of the RHS but not of the LHS of (F_3) . "Get broken," "are sat on," "are sawn into little bits," "are glued together," "have solid parts," "are of uniform material throughout," "contain no space" of the RHS but not of the LHS of (F_4) . "Has many members" of the RHS but not LHS of (F_7^1) and (F_8^1) . And so on.

(B)-hypotheses are heterogeneous identifications. "Contain a negative feed-back loop," "are detectable indirectly upon cutting open a cat's skull," "travel along nerve fibers," and "occur in brains," are significantly predicable of the LHS but not of the RHS of (B) and (B_1) ; and perhaps "hurt" and "are exhibited in external behavior" not of the LHS of (B) and (B_1) .

Not only (B)-hypotheses but identities like (F_8) and (F_9) and well-confirmed and accepted theoretical identifications like (F_1) and (F_2) , indeed all heterogeneous identifications, flounder when extensional identity is adopted as identity criterion. For how can a and b be identical if there is an extensional property f_0 such that " $f_0(a)$ " is significant but " $f_0(b)$ " is not? a and b would not have all their extensional properties in common, and so would fail to be identical even under the extensional criterion. A way to avoid these disasters, which has considerable appeal for other reasons as well, is to replace (4) by (7):

$$(7) x \cong y =_{\text{Df}} (\forall \text{ ext } f) . \Sigma^<f(x)> \& \Sigma^<f(y)> \supset . f(x) \equiv f(y);$$

to adopt what we call *contingent identity* or *coincidence* as defined by (7) as at least a *necessary* condition for the truth of heterogeneous identification statements. Call this condition "*requirement (a)*" on contingent heterogeneous identification. More generally, (6) is replaced by

$$(8) x \cong_R y =_{\text{Df}} (\forall f \in R) \Sigma^<f(x)> \& \Sigma^<f(y)> \supset . f(x) \equiv f(y),$$

where ' Σ ' symbolizes the predicate "is significant," a predicate whose significance-domain includes all quotation-expressions. In the theory of significance presupposed there are always predicates, e.g., "is an item," "is an occurrence," "is had by a ," "lasts

time t " in the case of (7), such that the antecedents of the definiens of (7) and (8) are satisfied.

When an identification is homogeneous (7) reduces to (4) if non-significant sentences are mapped uniformly into false statements, or else uniformly into true statements. A similar move would not rescue (4) in the case of heterogeneous identifications. If non-significant predications were mapped into true attributions then all identity statements like (F_8) and (F_9) (where negative significance-ranges of LH and RH expressions differ) would come out false; while if non-significant predications were mapped into false attributions all identity statements like (F_1) and (F_2) (where positive significance-ranges of LH and RH sides differ) would be swept away.

A favored manoeuvre, made to avoid replacement of (4) by (7) and sometimes to avoid replacement of (1) by (4), is to redefine refractory predicates so that they can be transferred from one side of an identity sentence to the other preserving truth-value. Suppose

$$(a) [x = y] \text{ is true, but } \Sigma^<f(x)> \text{ and } \sim \Sigma^<f(y)>.$$

Then ' f ' is so redefined and/or ' y ' is enriched by such conventions that at least

$$(b) [f(x) \equiv f(y)] \text{ is true.}$$

But, first, the manoeuvre presupposes (7). Allowing redefinition or the consistent addition of conventions here is tantamount to recognizing significance requirements. For if (7) is not correct the supposed case (a) is impossible: how, otherwise, could $[x = y]$ be true if predicates like ' f ' were not discounted? The same points tell against any attempt to save (1) by redefining modal predicates so that they transfer across contingent identity sentences. Allowing such redefinition amounts to discounting or dismissing the initial properties. Second, without (7) how is it to be decided which predicates can be non-viciously redefined? How can predicates whose redefinition is permissible be distinguished from predicates whose properties *falsify* the identity statement? Suppose that redefinitions of predicates (and referring expressions) are proposed without the introduction of (7). Then the redefinition ploy is vicious and very likely circular. To make appropriate redefinitions the identity should be known to be true. But if the identity is true redefinition would be otiose as far as guaranteeing the identity goes. If redefinition is needed to bolster up the identity then it must be in some way defective: how then can it form the

basis of appropriate redefinitions? Unless the identity statement is true and its truth determined independently using an identity criterion like (7) it cannot be made a suitable basis for redefinition. If defective identities could be made the basis of redefinition there would be serious danger both of rendering apparently correct defective identities and of guaranteeing the correctness of an identity only at the cost of its trivialization.

To stress these important matters consider the full materialism:

(M_1) : minds = brains

since it is much more vulnerable than (B)-hypotheses. For there are predicates like "are spatially locatable" and subpredicates of "are material things" like "are colored," "are not heavy" which are significantly predicable of the RHS of (M_1) , but what of the LHS? Redefinition of these predicates so that a "mind" has the location, color, and weight of its correlated brain not only begs the question since the truth of (M_1) is assumed in the definitional extensions of the predicates but also (what is in general inadmissible) changes the sense of "mind." If the sense of "mind" is sufficiently amended, in the limit to "brain," (M_1) will certainly be consistent: but trivially so. Worse, all the difficulties of extensionalizing moves are then encountered. In particular, the identity statement (M_1) is *not synthetic* if the sense of "mind" is sufficiently amended. Also, then, the Black dilemma and related puzzles are unavoidable.

A vital question, if redefinition and extension of significance ranges does not always trivialize an hypothesis, is: When is redefinition admissible? At least when (a) occurs and the significance range of ' f ' can (significantly) be extended as in (b) on the basis of the true identity $[x \cong y]$, where the truth of $[x \cong y]$ is assessed by using (7); but not in general. If correct, it follows that redefinition of predicates which are not impure simply to render an hypothesis consistent is inadmissible; for the truth, and hence consistency, of the hypothesis is assumed for redefinition. It also follows, since $[f(y) \supset \Sigma' f(y)']$, that the significance-range of ' f ' is extended under redefinition and that the sense of ' f ' is revised. Such (conceptual) revisions *do* occur on the basis of identity statements. Consider the extensions of significance-ranges of color predicates to include expressions referring to certain electromagnetic phenomena. Using our theory based on (7) these revisions can be comprehended.

Replacement of (4) by (7) at once eliminates those criticisms of (B) based on the existence of properties of physiological processes, the predicates of which properties are not significantly predicable of sensation-expressions, and mitigates those criticisms based on the presence of properties of sensations whose predicates are not significantly predicable of physiological referring expressions. To destroy this last criticism altogether requirement (β) on heterogeneous identifications [see below] has to be met. Roughly, certain emergent properties of sensations have to be explained in terms of the reducing physiological theory much as, in the case of (F_3) , macro-properties of material objects such as temperature and solidity (in the relevant sense of "solid") are explained under the reducing theory in terms of microstructures. In other words, replacement of (4) by (7) eliminates or softens several criticisms directed at (B)-hypotheses because they are heterogeneous identifications. The criticisms are thwarted because to establish contingent identity it is sufficient to show, for every extensional property f whose predicate is significantly predicable of SENGHS and of PROGS:

$[f(\text{SENGHS})] \equiv [f(\text{PROGS})]$.

Now restrict consideration to synthetic identifications; so mathematical examples can be discarded. Though there are grounds for distinguishing structural and state identities from process and event identities we shall not develop such distinctions. In fact, but not necessarily, all remaining identifications in (I) are observational identifications made from observational bases largely on grounds of spatio-temporal continuity and coincidence, non-separability, non-independent manipulability, and preservation of observational features. (Puzzles of personal identity and about sameness of boats and cars under various changes eject most criteria for sameness of actual referents in homogeneous identifications.) These grounds are also important in theoretical identifications falling under (II) but they are *not* nearly *sufficient*. For one thing, requisite observations may be physically impossible or seriously limited. The scientific identifications made under (II) exhibit a certain asymmetry. First, one referring expression in each, the one written typically on the right in English, is a theoretical expression explained (partly) within the context of a certain theory, while the other expression is usually an observational expression. Second, the identification is explanatorily directed; the item referred to by the

right expression is designedly introduced or used to explain the item referred to by the left expression and features of the item. This helps elucidate the point of the dictum that the "is" in heterogeneous identifications is the "is" of explanation. The dictum is misleading because there is not a special sense of "is" used just in heterogeneous identifications and because explanatory power is not a sufficient basis on which to make heterogeneous identifications. Moreover, if "Lightning phenomena can be explained by the theory of electrical discharges" were all that "Lightning is an electrical discharge" meant, we could still ask about the relation between lightning and electrical discharges. Because of the asymmetry we can distinguish LH and RH expressions as *reducenda* and *reducentes* respectively. The identification sentences are *reduction sentences* and the identity expressed in each is the *reduction relation*. A *heterogenous reduction* is a reduction where there are predicates significantly predicable of the *reducendum* which are not significantly predicable of the *reducens*; or, where a whole theory or a region of discourse and truths and well-founded or accepted assumptions and discriminations characteristically expressible in its vocabulary (a conceptual framework?) is being reduced, a reduction where the reduced theory includes in its formulation or vocabulary distinctive (descriptive) predicates which are not included in (initial) formulations of the reducing science⁶. The class of heterogeneous reductions overlaps that of heterogeneous identifications. (B)-sentences formulate heterogeneous reductions and theoretical identifications. That many expressions referring to sensations, like "Tom's toothache," "My giddiness," are observation expressions, and that many of the requisite neurophysiological terms needed in the identifications will be theoretical, though both controversial, seem to us defensible claims.

That (a) is not a sufficient condition for a theoretical identification is in part a consequence of the requirement that the reduced phenomena (or items) and further properties of the reduced phenomena be explained by the reducing phenomena (and the theories in which it is treated). The *reducens*' referent cannot provide the requisite explanation on its own. The identification is not made independently of the theory to which the *reducens* belongs. It is the availability of such a theory, containing law statements in terms of which

the reduced phenomena can be explained and their further related properties explained or predicted (perhaps using also associated theories), that provides the *further* grounds or case for the identification. If it were not for the theoretical connections the (open) universal statements yielded by sentences of (II) could not be adequately tested or confirmed. The further case for adding an explanatory requirement is this: With theoretical identifications several of the usual tests for identity, those used in observational identifications, are ineffectual or inapplicable; further the class of putative common properties is considerably reduced. To compensate for this diminution of requirements and to guarantee the dependence and concomitant variation of the identified items some further requirement is needed. A suitable requirement, which preserves the asymmetry of heterogeneous identifications, is the explanatory requirement. Lastly, in order to decide between rival verificationally equivalent hypotheses [see Sect. IV] further conditions on identity are needed.

An explanatory requirement (β) constitutes a necessary condition for the truth of a theoretical identification. Let ' x ' range over *reducenda*, and ' x_0 ', ' x_1 ' . . . be *reducenda*, e.g., expression referring to sensations. Let ' y ' range over *reducentes*, and ' y_0 ', ' y_1 ' . . . be *reducentes*, i.e., expressions referring to reducing items. Let the identity be formulated, generally, " $w \cong v$ ", where ' w ' designates the class of reduced items, i.e., the class of x 's and ' v ' the class of y 's. Then we require:

(β): There are scientific theories T_1 and T_2 such that y -expressions belong to the vocabulary of T_1 , and such that T_1 and T_2 —possibly together with experimental data about y 's and other items treated in T_1 —jointly explain both x 's and extensional (T_1 & T_2)-linkable properties of x 's whose predicates are significantly predicable of y -expressions, and also explain some extensional (T_1 & T_2)-linkable properties of x 's whose predicates are significantly predicable of x -expressions but not of y -expressions. A property f is *T-linkable*, i.e., theory-linkable under theory T , if f is not a relational property relating those items of which it is a property to items whose attributed behavior is not within the scope of T . For example, frightening many females is a relational property of lightning but not a property we expect electromagnetic theory to explain.

To illustrate: in order that hypotheses (B) be

⁶ See E. Nagel, *The Structure of Science* (London, Routledge & Kegan Paul, 1961), p. 342. The "homogeneous-heterogeneous" terminology has as sources the cited text and the work of B. Russell and of L. Goddard on significance-theory.

established there must be formulated a theory P_1 of physiological processes which together with more comprehensive theories P_2 , e.g., chemical and electromagnetic theories and possibly quantum theory, explains all those extensional (P_1 & P_2)-linkable properties of sensations whose predicates are significantly predicable of expressions referring to physiological processes, e.g., those relating to duration, pattern, intensity, location of sensations, and most (ultimately all) of those genuine extensional (P_1 & P_2)-linkable properties of sensations whose predicates are not significantly predicable of physiological processes.

(β) is a powerful requirement. The truth of statement (β) entails that all the standard *requirements on scientific explanations are satisfied*,⁷ which in turn entails that the usual requirements on a scientific reduction, insofar as they are applicable to sensations (since we do not have a scientific theory of sensations) are satisfied.

Finally (α) and (β) are jointly sufficient conditions for heterogeneous theoretical identifications or reductions like (B). If only a very limited class of sensations, say tickles, were identified with physiological processes then it would be a fair criticism that it is also necessary that further items, e.g., further features of *other* sensations be explained as well as those of tickles. When, however, occurrences as varied as the various sorts of sensations are explained and connected, the further necessary condition envisaged is automatically met.

It seems to us that (B)-statements, or at least statements yielded by (B)-sentences when the variable-places are filled in, can *consistently* satisfy requirements (α) and (β). We can, in fact, already go some little distance toward outlining general features of a requisite theory P_1 . Whether, however, such a theory when developed will contain or permit the expected reduction of sensations, seems at present open. If our surmises are correct (B)-hypotheses are *viable but only hypotheses*.

III

A major objection to combat is: (B)-statements cannot consistently satisfy a contingent identity criterion. Most discussion of (B)-type hypotheses

has, in effect, centered either on whether this major objection is correct or on whether (B)-hypotheses are empirical and on means of deciding between (B)- and allegedly rival hypotheses. The second of these issues we reserve for (IV) below; to the first we now turn.

It is argued: there are predicates, satisfying the antecedent of (7), such that the properties they specify are properties of SENGHS but not of PROGS; most notably the pairs "public," "private" and "corrigible," "incorrigible." We have pointed out that the properties specified by "is incorrigible to a " and "is about something private" are non-extensional, and thereby excluded. Suppose, in reply, exception is taken to the extensionality restriction itself, that it is objected to as being much too restrictive, that it is contended that the class of properties admitted under (7) should be widened to include other properties like incorrigibility. Arguments independent of (7) can be advanced for excluding such properties; e.g., comparisons of giddiness zones and pain patches with arid zones and brown patches can be used in arguing that being private and incorrigible are different *sorts* of properties of sensations from lasting t seconds or regularly exhibiting an α -rhythm and are not relevant to scientific identifications. If, however, we were to play along with the opposition the way would be open for the launching of an argument like the following:⁸

1. [p is a first person sensation report (statement)] implies [p is an introspective report]
2. [p is an introspective report] implies [p is wholly about something private]
- ∴ 3. [p is a first person sensation report] implies [p is wholly about something private]
4. [p is wholly about something private] is incompatible (inconsistent) with [p is about something public]
5. [p is a physiological (physical) report (statement)] implies [p is about something public]
- ∴ 6. [p is a first person sensation report] is incompatible (inconsistent) with [p is a physiological report];

from 3, 4, 5 using propositional calculus (or modal logic).

⁷ On these requirements, so far as they are worked out at present, see C. G. Hempel and P. Oppenheim "Studies in the Logic of Explanation," *Philosophy of Science*, vol. 15 (1948), p. 135; C. G. Hempel, "Deductive Nomological vs. Statistical Explanation," *Minnesota Studies in the Philosophy of Science* (ed. H. Feigl and G. Maxwell), vol. 3 (Minneapolis, University of Minnesota Press, 1962); and E. Nagel, *ibid.*, pp. 30-46.

⁸ This argument and the set of logically necessary conditions for *privacy* are extracted from K. Baier, "Smart on Sensations," *Australasian Journal of Philosophy*, vol. 40 (1962), pp. 57-70. The argument is a reconstruction of Baier's main argument against Smart.

The argument is valid but several premisses are either suspect or defective. A combination of the following points is sufficient to destroy the argument: First, it is misleading to call first person sensation reports "physiological reports," because an identity of sense is suggested. The employment of the indeterminate word "about" does not help in rectifying this fault. In the requisite sense of "about," in which two different expressions are *about* the same thing and where contingent but not strict identity obtains, ' x_0 ' can be both about something private (a sensation) and about something public (a physiological process). Second, premisses 2, 3, and 4 are false; they depend on a neat but questionable public/wholly private dichotomy. The usual senses of "public" and "private" (v. OED) cannot be used to support the dichotomy or in defence of 3, 4, and 5. Thus to back the argument technical senses of "about something private" and "is private" must be introduced. Consider the following fairly typical dualistic set of logically necessary conditions (with a brief justification added parenthetically) for the truth of " $f(x_0)$ is about something private" or " x_0 is private":

- (a) x_0 is owned, i.e., is had by someone.
- (b) x_0 is exclusive or unsharable (because two cannot share *one* pain).
- (c) x_0 is imperceptible by the senses (for "a sensed [e.g., saw, felt, tasted] a's pain," "a sensed b's pain" are both non-significant).
- (d) x_0 is asymmetrical with respect to persons (for "I could sense [e.g., see, hear] I had a pain" is non-significant but "I could sense he had a pain" is significant; and x_0 is something about the having of which the haver could not, but other people could, significantly consider or weigh evidence; e.g., "I am weighing the evidence for my being really in pain" is non-significant).
- (e) x_0 is such that the person who has it has final epistemological authority about it (for "I have a pain unless I am mistaken" is non-significant. The authority is based on the haver's being the person best placed to discover his mistake or confirm his belief).

These conditions are not only an inadequate base from which to spring to premisses like 3, 4, 5, but indeed unsatisfactory on other counts. First, observe that there are types or sorts of sensations as well as token-sensations. A type/token distinction can be readily transferred or applied to processes and occurrences. In fact many physiological processes and a good many sensations have been typed or classified. Token-sensations are (almost)

always had, experienced, undergone by an *individual* person, they can be assigned a given time bound, and they do not recur even though they may be a recurring type of sensation and have recurrence patterns within the time bound. Often, however, we refer to type-sensations; then a person may have the same sensation, e.g., a certain rheumatic pain, at widely repeated intervals, and several persons may have the same sensation, e.g., the same after-imagining, the same sort of, or the same, thirst or giddiness. Now token-brain-processes satisfy both (a) and (b); for they belong to and are in fact had (even if "felt" is not significantly applicable) by one person and they are exclusive, except possibly in cases like Siamese twins which create special difficulties about token-sensations too. One effect of (b) is to restrict consideration to token-occurrences: type sensations are not private. Against the comeback that it is not logically necessary that token-brain processes are owned and exclusive two moves are feasible: either to argue that this doesn't matter because (B)-statements are only synthetic; or to query the comeback and point out that any argument to this effect can be matched by an analogous argument in which "token-sensation" replaces "token-brain-process." Next, (c) seems to be mistaken unless qualified; pain-receptors play an important part in the "perception" of pain, e.g., when hit with a cricket ball. Moreover "a felt b's pain" is significant, unless "token-pain" is so defined that a token-pain is unsharable (as suggested in (b)). Such a definition is unpersuasive: if a's g-nerve were connected to b's g-nerve, through which pulses correlated with b's brain were being transmitted by a machine which did not interfere with pulses in b's nerve and which produced a pulse pattern (taking account of feedback effects if any) in a's nerve exactly similar to that in b's then we should say, probably along with most English speakers we guess, that a felt b's pain, i.e., one and the same pain. Incidentally, it can be admitted that one's own pain-experiences are perceptible in a much more direct way than any physiological processes, and that they are perceptible in a more straightforward way than another's pain-experiences without damage to identity theses. And, relying on (7), it can be conceded that while, e.g. "a feels his own pains" is significant "a feels his own brain processes" is not. The cogency of condition (d) depends on how "asymmetrical" is elaborated; for being sweaty under the arms, or having a mouth ulcer, is asymmetrical but not

wholly private. Elaborations suggested by the supporting examples should be jettisoned; for "I felt I had a pain," "I thought I had a pain," "I am debating whether I really have a pain" (I have a toothache) when the putative pain is of low intensity, and the seldom necessary "I can jolly well see I'm in pain, not just feel it" when remonstrating with someone, are significant, and the statements yielded may be true. Also a person may well have to weigh evidence during a sensation to determine what sort of sensation it is, or after an occurrence during which the weighing of evidence is out of question. Finally, it is not analytic that a person is the final epistemological authority on his sensations: (e) is too strong if genuine pains are to rate as *private*. A person is only the final authority on when he thinks or sincerely supposes he has a sensation; and having a sensation or having a genuine sensation *differs* from thinking that one has a sensation, and this from sincerely reporting that one seems to have or thinks one has a sensation. He is the final authority on when he thinks he has one because "if *a* sincerely supposes he has a pain then he has a supposed, or imagined, pain" is analytic. Admittedly a person is the authority on his own pains, because he feels them. Admittedly "*a*, in general, knows better than others when he is in pain" is true: only the statement is not analytic, and the "in general" cannot be elided. When neurophysiological theory is much better developed than at present we expect that public and physiological tests for sensations will be stressed much more; and that in some cases the results of these tests will override a person's reports. Compare the evolution of different tests for the comparison of temperatures and weights.⁹

Since pain-occurrences present as tough a front as any sensations in respect to privacy and incorrigibility let us, in our further defense, concentrate on them. To the questions "Does the sentence 'If I feel a pain then I have a pain' sometimes yield a synthetic statement?" and "Is there room for the expression 'I seem to feel a pain' as distinct from 'I do feel a pain'?" we propose to answer yes. We submit that the English word "pain" stretches, or ranges, *between two contrasts*, first a (seeming) pain/no (seeming) pain at all contrast, and second a real pain/imaginary or imagined pain contrast. The contrasts dovetail for (seeming) pains are

either real or imagined pains. Though the "real pain," "imaginary pain" terminology occurs in several English idiolects,¹⁰ the basis of the distinction is not so clear. Often, however, the distinction seems to be made on physiological grounds; consider real toothache, genuine pain. We propose to adopt physiological grounds as necessary conditions for real or genuine sensations; a person has a real sensation, an *r-sensation*, only if he is in the appropriate physiological state. Appropriate state: for if an identity hypothesis proves correct imaginary pains are coupled or identifiable with *some* neurophysiological basis. What people seriously say will have a good deal to do, initially at least, with which physiological states are appropriate. We call the pains of the larger class, seeming or sensed pains, which are contingently connected with sincere reports, *s-pains*. It follows from the distinctions that feeling an *s-pain*, or seeming to feel a pain, is not the only criterion for actually feeling or having a pain, or for having an *r-pain*. Consider especially low intensity *s-pains* where one may check for tissue damage, tooth decay, etc. Or suppose *a* reports that he has a pain in his foot. Then *a* has an *r-pain* if he has the appropriate pain in his foot, a matter which can be checked, ideally, by two coupled dolor-meters, one on the appropriate nerves in his foot or toward the extremity of the relevant nerve in the case of amputations, and the other connected to probes in the appropriate area in *a*'s pain center. He has only an *s-pain*, if an identity hypothesis proves correct, if a pain-detector scanning neural sensation-areas for pain and sensations like pains responds; and his report is sincere only if as a matter of fact the scanning detector responds. Given the distinction between *r-pains* and *s-pains* proposed answers to the initial questions can be vindicated; for "If I have an *s-pain* then I have an *r-pain*" is not analytic since the statement is sometimes false; and "I feel a real pain" is distinct from "I seem to feel a pain," i.e., from "I feel an *s-pain*." Contingent identity hypotheses are, however, *consistent* with other answers to the initial questions than those we proposed. For if such hypotheses are correct *situations will not in fact arise* where relevant physiological data and sincere sensation reports conflict, even though it is logically possible that such situations occur. If a repeatable situation were to arise where someone

⁹ See also the weight analogy in J. Wisdom, *Other Minds* (Oxford, Basil Blackwell, 1956), pp. 101-105.

¹⁰ An empirical claim which we have so far only checked over a small sample. Were it not for the apparent ubiquitousness of "real," "imaginary" terminology we should have chosen less controversial and ambiguous expressions; for "real" can mean, in these contexts, "acute."

felt a pain but the pain-meter didn't move though it seemed to be working properly, then we should have to abandon or modify the theory after sufficient rechecking. In this respect there is not more difficulty in testing the identity hypothesis, *once* we have the appropriate and calibrated instrumentation [within the limits of verification equivalence; see IV], than with most scientific theories—provided the hypothesis is openly falsifiable. The hypothesis may, however, at a later stage, having passed initial and repeated testing, cease to be openly falsifiable; it may become normic, i.e., nearer-analytic, and be adopted as *one* criterion for, or strongly-weighted mark of, a sensation. Physiological grounds would then be adopted as one criterion, and might even override apparently sincere reports of the person reporting that he has, or does not have, a given sensation. We might, under some circumstances, refuse to say that a person apparently sincerely reporting an after-image of a certain shape, size, and color was reporting correctly (if his visual apparatus, etc., was normal) if his report conflicted with the theory's predictions in respect to shape, size, or color. But if occurrences of this sort were frequent and varied enough we should be led to modify the theory. Even if such cases were to occur it would *not* follow that the sense of "I have a pain," etc., *must* have changed or shifted away from usual senses. For cases even now occur when a person may sincerely report a pain but many English speakers, at least, are prepared to say that the person did not *really* feel a pain, that he only *imagined* or thought he had a pain, or that he had an imaginary pain.

We outline documented support for the real/imaginary pain distinction. On October 23, 1963, in Sydney a woman was awarded Workers' Compensation by the full Supreme Court on the strength of delusional pains she had which were produced or contributed to by an accident at work. Both the judges and the ABC news reporter used the expressions "imaginary pains" and "delusional pains"; and the ground on which the pains were so-called was that there was no physiological basis for such pains.¹¹ It could be argued, though none too convincingly, that the expressions simply meant that she had no obvious damage or diseases rather than that there were no appropriate physiological processes or that she wasn't in real pain. None-

theless the case certainly seems to show that there might be situations where it is of interest and importance to use physiological criteria for *genuine* pains.

Many arguments alleging that (B) cannot consistently satisfy a contingent identity criterion lean heavily on the supposed incorrigibility of first person pain reports. But such reports are not absolutely incorrigible (to set out points summarily) because (i) a person can be mistaken and corrected as to whether he has *r*-pains. If we consider from now on only *s*-pains, then indeed a person can't be corrected if he *succeeds* in sincerely making the statement "I have an *s*-pain"; but only because if he sincerely reports that he has a pain, then he has an *s*-pain. And his sincerity may be questioned. Moreover, (ii) a person's *s*-pain statements may be misdescriptions. It is a mistake to dismiss misdescription and misclassification along with slips of the tongue, etc., as simply verbal errors. For in misdescription these errors are reflected in wrong or false statements, statements which might have quite serious non-verbal consequences. And the misdescription may result from lack of knowledge or mistaken beliefs about the pattern or features of his sensation; in which case it is not merely his words but his *beliefs* which are incorrect. Under the same heading we may include mistaken comparisons, which may occur, because of slips or, more important, because of mistaken assumptions or beliefs about language or about non-verbal matters. (iii) "Incorrigible" associates with "correct" and "incorrect." These words have multiple roles: "incorrect," for example, is highly determinable and covers a wide variety of criticisms; consider the sorts of items which may be called "incorrect" and the various different reasons for the incorrectness. Part of the guile of the incorrigibilist is not to disclose which reasons or criticisms he is prepared to take into account as relevant. "Incorrigible" can associate with "impossible"; incorrigibility should be concerned with the *logical* impossibility of correction of the speaker. In less stringent senses of "incorrigible" first-person pain reports are virtually incorrigible, e.g., there is little *real* possibility that the speaker will be corrected. (iv) The endeavor to rule out unfavorable examples by progressively tightening "incorrigible" so as to eliminate, along with lying and

¹¹ Two key passages read as follows: (i) (Mr. Justice Moffit in his judgment) "It was accepted as a fact that the applicant has a delusional state causing her to experience the symptoms of pain referred to," and (ii) (Sydney *Morning Herald's* Law Reports, October 24, 1963): "... this psychosis had produced after the accident a delusional state giving the applicant an impression that she was suffering such pain and discomfort in the right side that she was unable to perform her work."

slips of tongue, misdescription mistaken comparisons, and mistakes of language, hits other snags. For what is it he can't be *corrected* with regard to? It can't be his statement: it must be his intended statement or his belief. Now if he is at all specific he can still be mistaken; for he may believe that his sensation of sort g_0 is of sort g_1 , or that he had a sensation when in fact he had something else a bit like a sensation. The incorrigibility is only preserved by making his intended statement or his belief sufficiently *vague*. This can be done not only with sensation statements (e.g., "I had now a *something* or other"), but with other sorts of statements as well (e.g., "Something has happened to me now"). Incorrigibility and vagueness do not vary independently. It seems that as we increase incorrigibility we increase vagueness, and that in the limit an absolutely incorrigible "report" or "belief" is absolutely vague—not really a description, report, or belief at all. (v) A set of non-equivalent definitions of "a's statement '*p*' is incorrigible," definitions based on criteria extracted from various authors, reveals that the exact sense of "incorrigible statement" lacks clarity. This point is argued in detail in the Appendix.

Other contrasts than public/private and corrigible/incorrigible contrasts, such as physical/mental and observable/theoretical contrasts may appear to block contingent identity hypotheses. They don't. If these predicates marked out properties which created difficulties for heterogeneous identifications similar difficulties would arise in the case of accepted scientific identities. The hypothesis (F_1) that temperature is mean kinetic energy does not run into difficulties over thermodynamical/mechanical and observable/theoretical contrasts: similarly (B)-hypotheses do not strike problems with these other contrasts. Subject-classifying and category predicates like "electrodynamical," "mental," and "subjective" do not specify properties which succeed in destroying contingent heterogeneous identifications.

On the contrary (β), if fulfilled, provides a condition under which a subject can be extended. Then objections based on classificatory predicates vanish because the classificatory properties cease to be exclusive. For example, with the advent of the reduction of temperature under the kinetic theory a phenomenon could correctly be classed as both thermodynamical and mechanical. Similarly if (α) and (β) are satisfied in the case of the identification of SENGHS and PROGS mental and physical are not exclusive. Indeed given (α) and

(β) objections based on classificatory predicates could be crushed simply by reclassification. Classifications are fluid: they change with contingent identifications, which destroy the exclusiveness of classifications. Nor are subjects as sharply separated as the objection would suppose. "Mental" and "physical" for instance do not mark out exclusive properties: various activities such as driving with care and reading are correctly classed as both mental and physical; and sensations are investigated in physiology. Next, subject classifications depend on the state of our knowledge and beliefs. Therefore classificatory properties which directly reflect this dependence are non-extensional. But contingent heterogeneous identifications are unaffected by variance of intensional or impure properties. For this reason too, observational/theoretical contrasts provide no basis for criticism of contingent identity claims: for the relevant properties are not extensional. Consider, to demonstrate this point, material equivalences between statements like "That is a violet light" and "That is electromagnetic radiation of wavelength approximately 400 Å." The predicates "is an observational statement" or "That . . . is observational" and "is a theoretical statement" do not transfer across the equivalence preserving truth-value. If, contrary to our view, one of these predicates were always to transfer across such material equivalences the observational/theoretical contrast would not prove an obstacle to contingent identifications.

Finally, although sensations generally have locations, e.g. the pain is in his foot, her head is giddy, a location problem still threatens (B)-hypotheses. For though a man's pain may be in his foot relevant physiological occurrences may occur all the way from his foot to pain centers in his brain. But if one location is less extensive than the other how can there be contingent identity? First, the *felt* location of a given pain is an intensional property of the pain, and so is discarded for contingent identity. According to OED ". . . is in —" in the relevant sense means "the part affected by . . . is —." This definition needs revision to allow for chiropractors' reports and for the man with an amputated leg who insists that the pain is in his foot. It seems that "*a*'s pain is in —" means "the part affected (the painful part) is felt (thought) by *a* to be —." The predicate "the part affected by . . . is felt by *a* to be *h*" demarcates an intensional property; therefore the synonymous predicate "*a*'s . . . is in *h*" also does. Second,

some physiological processes do appear to have felt locations as well as actual locations, e.g., those identified with pains at the extremities of the nerve fibers in which the processes occur; and *felt locations* could be suitably defined for some of those for which predication of a felt location would be non-significant.

IV

(B)-hypotheses are no more than contingent identities. Merely contingent identity statements are *contingent* statements. For if such a statement were not contingent but analytic, then the identity would hold for some modal properties; and then the identity would not be merely contingent. Since the restricted identity theory is merely contingent there is no reason why the terminology of the theory should conflict with rather than refine and complement everyday terminology for discoursing about sensations or the theory, if correct, with the truth of many everyday statements about sensations. Even if the identification were adopted as a criterion, e.g., of real sensations, it would not follow that everyday discourse must be abandoned and replaced by terminology drawn from the theory. The fact that in some circumstances we use occurrences of electrical discharges as a criterion for whether or not lightning occurs does not prevent or destroy phenomenal or ordinary descriptions of lightning, nor imply that we have to abandon such discourse and associated tests and speak instead in terms drawn from the theory. Similarly even if situations sometimes arose where we were to use physiological criteria to decide whether someone really felt a pain, these would not prevent us making the statements we do nor force us to withdraw or reject as false many of them. On the contrary, since "... occurs" and "... exists" specify extensional properties, it follows if identity (B) is true that SENGHS occur and exist if and only if PROGS do. *Nor*, in general, does the introduction or establishment of contingent theoretical identifications *change the sense* of the observational *reducentia*. Just as there was, with "red," "hot," "lightning," though the senses of the words were determinate, room for further theoretical connections and identification, without change (or substantial change) of the sense of the words, so there is with sensation expressions room for heterogeneous identifications without change of sense of the expressions. In this respect an observational

expression ' x ' is *theoretically completable* without changing its sense, even though tests or criteria for x 's may shift once the theory becomes well-established and identity statements are normic. Introduction of normic connections, in place of merely contingent connections where the connections are well-established, is possible with cluster and multi-criterion words without change of sense of the words. Just this happens along with the development of theories. As in the lightning example, so with sensations there would be justification for introducing theoretical connections. There is not only room, but a need for them.

Subject to certain qualifications, e.g., relative to an embedding scientific theory P_1 given which the unspecified constants in representative sentences are specified, (B)-statements are *empirical*. A modified verification criterion can be used to establish this point. For example, (B)-statements are empirical because together with empirical statements about SENGHS they (minimally) entail empirical statements about PROGS. (Compare tests for r -pains indicated in Sect. III.) But it does not follow, just because (B)-hypotheses are empirical, that there are not other hypotheses which are verificationally indistinguishable from them. Notwithstanding, it is frequently assumed that the *maximal empirical statement* which evidence could support can be formulated something like (J): Any sensation of sort $h(g_0)$ in x is correlated one-to-one with and is (approximately) simultaneous with a physiological occurrence of sort g_0 in x . That is, there is a class (presumably finite) of sensations of sort $h(g_0)$ in x , and a class of physiological occurrences of sort g_0 in x , and these classes are correlated by a one-to-one relation, say R_0 . Further, each item, PROG or SENGH, occurs at approximately the same time as its correlate under R_0 , in the same time-slice Δt , say.

Difficulties infect the obtaining of a non-trivial correspondence because the occurrences involved are extensive, complex, and hard to distinguish from one another except in an arbitrary way, or, in the case of PROGS, in a way *just* based on patients' reports of SENGHS. To highlight these difficulties consider the following. We assume first that at each time t , x_0 has a SENGH there is some physiological occurrence in x_0 , second that x_0 only has one SENGH at a time, and thus within bounds Δt containing t . Then we can set up a trivial correspondence satisfying (J) by defining as the correlate of SENGH at t all physiological occurrences in x_0 in interval Δt . To avoid such difficulties the assumptions,

especially the second, must be dropped, much more detailed descriptions and specifications of SENGHS and PROGS and their patterns must be given, and the occurrence and extent of PROGS must, after initial calibration of instruments and determination of extent, be marked off independently of the occurrence and extent of SENGHS. Restriction to a single person x_0 should be dropped. Compare, as regards calibration and the setting up of correspondences, the beginnings of thermometry.

The thesis analogous to (J) for lightning runs: (H): Any lightning (flash) is correlated one-to-one with and is s.t. (i.e. spatio-temporally) coincident with an electric discharge (of sort) d (i.e., from an ionized cloud to earth, etc.). Here and below, "(s.)t. coincident" abbreviates "approximately (spatio-)temporally coincident." On the basis of (H) together with electromagnetic theory several facts about lightning can be explained and predicted. For example, given initial conditions whether the lightning will be sheet, forked, or chain, its striking point can be predicted. So given electromagnetic theory there are good reasons for advancing beyond (H). But to grant (H) on its own is by no means to grant (F_2): also, as required under (β), further features of lightning must be explained and predicted.

Traditional one-track mind-body theories, seen in miniature in the sensation case, can be reflected in the lightning example. For the following hypotheses might be put up on the strength of (H).

- (H₁) Between the *different* items lightning flashes and the occurrence of electrical discharges d there is a mere 1-1 correspondence and s.t. coincidence; and since respective correlates are different they are two.
- (H₂) Lightning is caused by (and s.t. coincident with) electrical discharges d (but not conversely).
- (H₃) Between (the different items) lightning and electrical discharges d there is two way interaction (and s.t. coincidence).
- (H₄) Lightning and electrical discharges d are just two (corresponding and s.t. coincident) aspects of the one underlying process.
- (H₅) (F_2), which implies that there is only *one* item.
- (H₆) Really there is no such (separate) thing as "lightning" but only electrical discharges d .
- (H₇) Lightning is nothing over and above electrical discharges d .

The bracketed expressions are included in an hypothesis just in case they are not rendered unnecessary by the remainder of the hypothesis.

Further hypotheses, e.g., single aspect, could be added.

A corresponding set of hypotheses (J_1)-(J_7) can be obtained for the sensations case (J), by replacing "lightning (flash)" and "electrical discharge d " in (H_1)-(H_7) by "SENGH" and "PROG" respectively. Comparing the two sets of hypotheses in detail is valuable. It is easy to envisage situations in which the dispute over (H)-(H_7) might be just as alive as disputes over (J)-(J_7), e.g., if our religion or cultural-history included lightning-worship (not science-worship) and people were still aesthetically interested in lightning.

Given suitable expansion or explication of such key expressions as "cause" in (H_2), "interaction" in (H_3), and "aspect" in (H_4) and given more explicit formulation of certain hypotheses, member statements of a set of hypotheses are verificationally equivalent, as (H_1) is verificationally equivalent to each member of the (H)-set. Roughly, statement S_1 *verificationally-implies* (v -implies) S_2 when S_2 has no observational consequences which S_1 does not have, i.e., when S_1 and S_2 entail inconsistent (direct) observational statements. Let P be a given consistent class of statements; e.g., the statements of theory P_1 . Then S_1 v -implies S_2 w.r.t. P (i.e., with respect to P) iff all observational consequences of S_2 and P are observational consequences of S_1 and P' , where P' is a modification of P obtained by at most uniformly replacing RST (equivalence) relations of a well-defined class of sentences of P by different RST relations (and adding analytic statements). S_1 is v -equivalent to S_2 w.r.t. P iff S_1 implies S_2 w.r.t. P where P' is the modification of P and S_2 v -implies S_1 w.r.t. P' where P is a suitable modification of P' i.e., if all observational consequences of S_2 and P are observational consequences of S_1 and P' and conversely.

There are two stages in establishing the required v -equivalences. In the first the (J)-hypotheses are shown to divide into two classes, the (J_1)-class and the (J_6)-class, such that each hypothesis of a class is v -equivalent to every hypothesis of its class. But there are indeterminacies in hypotheses of the (J_1)-class which can be resolved in two directions, either so that (J_1)-hypotheses differ verificationally from (J_6)-hypotheses or so that the hypotheses are v -equivalent. Thus in the second stage we replace (J_1)-hypotheses by (J_1^+)-hypotheses, and explain why (J_1^+) is v -equivalent (w.r.t. the embedding theory) to (J_6). To accomplish the first stage preliminary results are useful.

1. If S_1 entails S_2 then S_1 v -implies S_2 w.r.t. any P . For if S_1 entails S_2 all consequences of S_2 are consequences of S_1 ; therefore all observational consequences of S_2 are observational consequences of S_1 .
2. If S_1 and S_2 are mutually entailing, or strictly equivalent, S_1 is v -equivalent to S_2 w.r.t. any P .

The second part follows because differences between mutual entailment and strict equivalence show up at most in non-synthetic consequences, whereas all observational statements are synthetic.

3. If S_1 and S_2 are v -equivalent w.r.t. P and S_2 and S_3 are v -equivalent w.r.t. P' where P' is a suitable modification of P then S_1 and S_3 are v -equivalent w.r.t. P' where P' is a suitable modification of P (or P').

Because of 3 it suffices to establish a chain of v -equivalences.

To sharpen (J_1) we define mere correspondence: classes w and v *merely correspond* iff they correspond one-to-one, i.e., there is a one-to-one relation between them, but there are elements x of w and y of v such that for some extensional property f , ' $f(x)$ ' and ' $f(y)$ ' are significant but x has f and y lacks f ; then:

If w merely corresponds to v , w is not contingently identical with v , i.e., $w \not\equiv v$.

In the *sketch arguments for the first stage P* may be any set of statements consistent with the conclusions of Sect. I. Now x and y *interact* iff there are causal relations between x and y . Since then (J_2) entails (J_3) , (J_2) v -implies (J_3) . In fact hypotheses (J_2) and (J_3) differ at most as to the direction of the causal relation in some instances. But in the case of sensations, causal relations are from physiological processes to sensations (see Sect. I). Therefore (J_2) and (J_3) do not differ verificationally here, i.e., differences between epiphenomenalist and interactionist hypotheses only appear in larger contexts. In coping with (J_2) it simplifies matters and seems permissible to use a regularity analysis of "cause": roughly, where determinate causal relations are not observable v causes w iff there is a one-to-one (lawlike) relation R_0 correlating v with w such that if xR_0y then x does not succeed y . (J_2) splits into two cases. Either $w \cong v$ and (J_2) coincides with (J_5) ; or $w \not\cong v$ and (J_2) is v -equivalent to (J_1) . For then (J_2) entails (J_1) . (Alternatively use the distinguishing property "... causes y_1 "). Conversely to show (J_1) v -implies (J_2) use the fact that if y_1 is the correlate of x_1 under R_0 then if x_1 is t.coincident with y_1 then x_1 does not succeed y_1 , and use a distinguishing property. (J_4) also splits

into two cases. Either the aspects are identical with one another and so with the underlying process, and (J_4) coincides with (J_5) ; or the aspects differ, and (J_4) is v -equivalent to (J_1) . For use:

$$R^2 \in R \cdot R = \check{R} \equiv . (\exists S) . S \in Cls \rightarrow 1 \cdot R = S/\check{S}.^{12}$$

If (J^1) holds, there is a transitive, symmetric relation R_0 relating w and v , and therefore there is a many-one relation S relating w and v to some underlying item. Thus, short of further explanation of "aspect" to exclude the equation, w and v are S -aspects of an underlying item, i.e., (J_4) holds. Conversely if (J_4) holds w and v are related to an underlying process by aspect relations, which we can combine as a many-one relation S_0 . Then $R_0 = S_0/\check{S}_0$ is a relation having the required properties for (J_1) . Alternatively consider the relation product R_1 of aspect relations S_0 and T_0 , and make use of the fact that the aspects are corresponding, t.coincident, and different. (J_7) implies (J_5) . If x 's are nothing over and above y 's therefore x 's have no properties that y 's do not. All properties that y 's have in addition to these properties are properties like f whose predicates are not significantly predicable of *reducentes*. For otherwise not- f would be a property of x 's but not of y 's, contrary to hypothesis (J_7) . Therefore x 's and y 's have all extensional properties in common; and $w \cong v$. The converse is clear, provided by "nothing" is meant "no extensional properties." In (J_6) the qualifying word "separate" is crucial; otherwise (J_6) could be refuted given (J_5) . Consequently also (J_6) should not be formulated "... do not exist; only—exist"; a qualification like "independently" is needed in the first clause if v -equivalence is to be preserved. (J_5) implies (J_6) ; for if x and y are identical there are no extensional features separating them; therefore they are not separate. Conversely that x and y are not separate implies that they are coincident.

Two classes of hypotheses emerge, the one-item (J_5) -class and the two-item (J_1) -class. Because the formulation of (J_1) is insufficiently explicit, (J_1) and (J_5) are not clearly v -equivalent. V -equivalences of two-item hypotheses only follow because of similar inexplicitness in other hypotheses of the class. The salient point is that it is not made explicit whether or not x_1 displays a similar observational pattern to its correlate y_1 . Similarity has yet to be explained; but suppose, to illustrate, we discovered that in a time interval of occurrence

¹² A variant of 72.66 of A. N. Whitehead and B. Russell, *Principia Mathematica*, 2nd ed. (Cambridge, Cambridge University Press 1925). The symbolism for relations is drawn from this text.

t_1 to t_{1+1} a typical PROG (or y_1) had a quite different intensity graph from its correlated SENGH (or x_1), e.g., the positive cycle of a sine wave opposed to a saw wave. Such data might be sufficient to falsify (J_5) -hypotheses and/or to require modification of underpinning theory P_1 . But the data would not falsify or disconfirm (J_1) -hypotheses unless these hypotheses also implied that the observational patterns of correlates were isomorphic. To progress we shall have to sharpen (J_1) -hypotheses. Either (a) (J_1) -specifications should be such as to require that correlates are observationally isomorphic, or (b) such as to require that they are not always observationally isomorphic. In case (b), (J_1) is not v -equivalent to (J_5) because (J_1) together with the relevant theory implies that there are correlates which have differing observational features, whereas (J_5) implies that there are not. Those who have espoused (J_1) -hypotheses do not seem prepared to commit themselves to (b); nor should they, unless they think that there is mainly an empirical difference between (J_1) and (J_5) , and that there is some evidence in favor of (b). So let us recast (J_1) -hypotheses so that they conform to (a). We abbreviate "observational" by OBS, and "observationally isomorphic" by *o-isomorphic*. Then we introduce a definition designed to capture essentially what we mean by "having identical observational features":

x is *o-isomorphic* with $y \equiv_{\text{DF}}$
 $(\forall \text{ EXT \& OBS } f). [\Sigma^* f(x)^* \& \Sigma^* f(y)^*] \supset [f(x) \equiv f(y)]$.

Next we replace (J_1) – (J_4) by (J_1^+) – (J_4^+) , where (J^+) is obtained from (J^-) by adding to the formulation of (J^-) :

"& $(\forall y, x)(yR_0x) \supset [y \text{ is } o\text{-isomorphic with } x]$."

In arguing for the v -equivalence of (J_1^+) and (J_5) we depend on the point that (J_1^+) and (J_5) are not themselves purely observational statements. Otherwise, in virtue of preceding definitions, they could not be v -equivalent. It follows that (J_1^+) and (J_5) differ (extensionally) only with respect to non-observational properties.

The relevant differences between (J_1^+) and (J_5) will appear in their embedding theories. Let the complete embedding theory of (J_5) be P_5 , where P_5 includes theory of physiological processes P_1 and requisite logico-mathematical and auxiliary theories. Observational consequences of $P_5 + (J_5)$

are obtained by making use of initial observation conditions. The initial conditions will be the same irrespective of whether (J_1^+) or (J_5) is used, but the theories used to derive observational consequences will differ slightly according as (J_1^+) or (J_5) is used. We want to show that (J_1^+) is v -equivalent to (J_5) with respect to P_5 . Thus we need to show that (J_5) can be replaced by (J_1^+) and the embedding theory by P_5' without diminishing the class of observational consequences, and similarly under replacement of (J_1^+) and (J_5) w.r.t. P_5 . Like reduction sentences (J_1^+) and (J_5) are only needed at the beginning and end of any derivation of observational consequences because other mathematical transformations can be carried out within the respective embedding theories P_5' and P_5 without making use of *reducenda* expressions at all. Because of results like: $w \text{ sm } v \& v = u \supset . w \text{ sm } u$, where 'sm' abbreviates "is similar to," correspondences will be preserved. Reformulate derivations of observational consequences so that all applications of (J_1^+) , or of (J_5) , occur at the beginning or end of respective derivations. There will be at most a finite number of applications of (J_1^+) , or (J_5) , at the beginning and end of any such derivation. Such transformations of proofs can be carried out effectively; thus a principle of substitution of identicals for *reducenda* with *reducetes*, the main technical advantage of identity relations over other equivalence relations, is not required.¹³ Suppose, then, in a derivation of an observational consequence using $(J_1^+) + P_5$ we have rearranged the derivation so that (J_1^+) occurs at only initial and/or final stages, and all intermediate steps are carried out without using *reducenda* expressions. We may in rearranging have to make use of simple equivalences from the theory of relations. Next, systematically replace one-to-one correspondence relations appearing in (J_1^+) connections by contingent identity relations, and thus (J_1^+) by (J_5) and P_5' by P_5 . Derivations of observational consequences are not destroyed, for a derivation could only be altered at initial or final stages if a property serving to distinguish mere correspondence from identity appeared essentially in the correlation. This could not happen at the end since the correlation is here employed in reaching observational consequences; so that there would have to be observational properties with respect to which (J_1^+) differed from (J_5) , contrary to our previous finding.

¹³ In fact a principle of substitution of identicals is all that is needed in first order functional calculi with identity to distinguish identity from other equivalence relations; and in the simple calculus of equality identity and equivalence relations coincide. See, e.g., A. Church, *Introduction to Mathematical Logic*, vol. 1 (Princeton, Princeton University Press, 1956), p. 281.

And the derivation could only be upset at the beginning if a distinguishing property were essential in getting from initial observational statements, or in reaching observational consequences. But then it would be an observational property contrary both to the requirement of *o*-isomorphism and to the neutrality of initial conditions. Similarly starting with derivations using $P_5 + (J_5)$ replacement of contingent identity relations in (J_5) connections by one-to-one correspondences preserving *o*-isomorphism, and thus replacement of (J_5) by (J_1^+) and P_5 by P_5' , does not diminish the class of observational consequences. This completes the sketch argument for *v*-equivalence.¹⁴

It is not true that correspondence hypotheses like (J) or (J_1) make fewest assumptions and approximate best to empirical data; and that therefore such hypotheses are all we are warranted in opting for. Unless revised as in (J_1^+) correspondence hypotheses may not even take account of all observational data. With hypotheses which are *v*-equivalent w.r.t. the requisite theory there is no basis for claiming that what has been established scientifically just is a particular one of them, e.g., a correspondence hypothesis, or that such a one has privileged status as "closest to the scientific facts." Such *v*-equivalent hypotheses have equally good empirical standing. The fact is that the (putative) scientific findings have *here*, in (H^+) and (J^+) , been presented in a way which, though fairly neutral (there are limitations on the inter-theory neutrality that can be got in the formulation of scientific hypotheses containing *theoretical* [not observational] expressions or expressions with important theoretical connections), is more favorable to some hypotheses than others, as to a correspondence than to an identity or double aspect thesis. For they suggest that there is mere correspondence and thus that there are *two* things which are correlated. Since this holds equally for

(H)-hypotheses as for (J)-, it is not true that to grant (H^+) is just to grant (H_5) . To grant a statement is not thereby to grant any statement *v*-equivalent to it w.r.t. accepted scientific theories, or there could hardly be so much philosophical rivalry. These factors should also emphasize difficulties for those who adhere to hypotheses of each set which do not correlate numerically, for example, to both (H_5) and (J_1) .

On what bases do we *choose* between (H_1^+) - (H_7) or (J_1^+) - (J_7) ? How do we decide what to *say* in (J)-cases where usage is not determinate or well-founded. Short of introducing, as with (a) and (β), sharper criteria, e.g., for whether *one* process or *two* processes occur, criteria which when accepted settle the issue, there is no satisfactory way of resolving the issue apart from the weighing of cumulative reasons for deciding in one way or on one hypothesis rather than another. And the defense of the introduction of such sharper criteria comes down to the weighing of such cumulative reasons. The issues are characteristic *conflict* or *decision issues* in which the outcome, if any, is not fixed by empirical matters to more than *v*-equivalence though empirical considerations may be important in reaching a reasoned verdict. There is a relevant difference between the (H) and (J) cases. The (H)-group at present provides us with an over-determined conflict case, i.e., the case has normally been resolved in favor of (H_5) and there is a fairly well-established usage. But the (J)-group presents us with an under-determined conflict case, where there are rival usages linked with different criteria and comparisons. This is not the only difference between the groups; for token-sensations are private and relatively incorrigible.

Serious conflict issues are not satisfactorily resolved either by chance (e.g., tossing a coin) or by applying Occam's razor or crude parsimony principles.¹⁵ To appreciate how more balanced

¹⁴ If *v*-equivalence were so defined as to require only sameness of observational consequences, Craig's theorem could be used to show *v*-equivalences.

¹⁵ Since we have been treated to large doses of these applications in recent literature we present briefly some defects in the views that one does, ought to, or must choose between hypotheses using Occam's razor (abbreviated OR). (a) OR does not discriminate between (H_4) and (H_5) and (H_7) , if aspects or appearances are not entities. (b) OR is not a panacea. If applied indiscriminately it could effect all sorts of reductions many of which have little plausibility. In order for it to be of any use it has to be heavily supplemented by further conditions; in particular, criteria have to be supplied for what is to count as an *entity*, on which issue OR usually begs the question, and for *necessity*. It suggests a pragmatic criterion for necessity, which is objectionable since the existence of entities is not a matter of their usefulness or practical necessity. (c) Unless these conditions are supplied OR looks remarkably like an unwarranted stipulation. We can ask for, but not seriously expect, a justification for OR. And there is nothing to stop us from setting up a rival principle. (d) OR is much too nominalistically oriented. For strictly we cannot either multiply or reduce most non-abstract entities or items; we can only as a rule multiply or reduce our terminology for talking about them. (e) OR is unnecessary. We can settle conflict issues, and handle "hypotheses" OR is often invoked to eliminate, without appealing to OR. So if a principle is an entity OR includes itself in its own scope. For related reasons parsimony principles have dubious value in resolving conflict issues unless carefully hedged.

decisions between rival hypotheses may be reached consider that major issue which turns on whether to say "one thing" or "two things," "one process" or "two processes." Deciding for identity of items as in (J_6) implies but is not tantamount to deciding in favor of "one item"; for if x_0 and y_0 are identical (not two things with identical properties apart from spatial or temporal properties) then x_0 and y_0 are one. If x_0 and y_0 only correspond then x_0 and y_0 do not have all properties in common, and are therefore distinct. Thus adoption of (J_1^+) or (H_1^+) implies deciding on "two distinct occurrences."

To appreciate reasons for saying "one occurrence" rather than "two occurrences" consider a series of cases. Consider first a furniture remover and his reasons for saying that we have two pictures but only one table, not both an Eddington table and a Stebbing table.¹⁶ Consider next the reactions of a lightning-remover expert to various of hypotheses (H^+)–(H_7). His client, an adherent of one of (H^+)–(H_3^+), says: "All right, that thing you put up gets rid of the electrical discharge and the crackle on the radio; but what are you going to do about the lightning and its burning effects? Don't we need some protection against that too?" In these circumstances we tend to seize on (H_4)–(H_7). None of (H^+)–(H_3^+) would normally be sufficient to allay the client's anxiety. We should have to explain somehow how we have complete control over the lightning once we have control over electrical discharges and that rendering one of them harmless renders the other ineffectual. Even a causal hypothesis like (H_2^+) doesn't surmount these difficulties at all easily. Connections between correlated items of (H^+)–(H_3^+) remain obscure, even mysterious, and insufficiently elucidated. Briefly, (H^+)–(H_3^+) are *explanatorily inadequate*. Now consider the case of someone who has a pathological fear of after-images, has convulsions when they occur, etc. We tell him about his retinal structure, optic nerve, and visual cortex and instruct him not to stare at bright objects for any length of time. He says, "That's all very well, but it's not my retina and whatnot I'm worried about—it's those weird visions I keep getting." Or consider an exponent of one of (J^+)–(J_3^+) who gets severe pains in his arms. His doctor gives him a pain-killer (-remover) like morphine and a rudimentary physiological explanation. But the patient says, "I understand how this stuff goes into my blood-stream, how it affects my neurons, but I'm

still at a loss to know how it will stop that dreadful stabbing feeling in my arm." How could we help him using his hypotheses? These hypotheses are unsatisfactory here because they make it look as if there is something further to be *explained*. They encourage questions like: *Why* and *how* does this correspondence occur? We can't answer this question without either blocking it with, e.g., "the correspondence is ultimate (unexplainable)" or "there are fundamental category differences (which are unexplained, unexplainable, or don't need explanation)," or else appealing to some further identity and connections. The first alternatives each create serious problems: some of them violate the principle (of sufficient reason) that every phenomenon has a (rational) explanation, and so isolate themselves from support; most give no account or explanation of the requisite ultimacy or fundamentality, and so on. With contingent identity these problems are avoided, since if (α) and (β) are satisfied, explanatory relations are coupled with identity relations. Since, however, the relevant explanations are mostly causal explanations (c.f., Sect. I, 8) it is imperative to allow, as we have, that " x causes y " is *consistent* with " $x \cong y$." Because it is quite fashionable to so characterize "cause" and "explanation" that explanatory relations are inconsistent with identity relations it needs stressing that such a course has devastating consequences; in particular that the making of heterogeneous identifications in science is inconsistent with the use of these same connections (in the way they regularly are used) to provide explanations, and that a full materialism like (M_1) is inconsistent with much of psychology.

Many other *reasons tell against* certain (H)- and (J)-hypotheses. Correspondence hypotheses, (H_1^+) & (J_1^+), as well as being explanatorily unsatisfactory, appear mysterious. Any mere correspondence can be destroyed without destroying the correlates; but an identity cannot significantly be destroyed, only the item can be. If a correspondence can be destroyed it makes sense to ask what keeps it going, what explains it, questions which are not significantly asked of an identity. *This* is why an identity succeeds in blocking requests for further explanation. For every equality and correspondence there is an underlying identity. For if v and w are correlated by one-to-one relation R_1 then correlated items are identical apart from certain extensional features which one has and the other lacks. If (J_1^+) were correct they would not be

¹⁶ See L. S. Stebbing, *Philosophy and the Physicists* (London, Methuen, 1937); and J. Wisdom, *op. cit.*, p. 67.

observational properties. Suppose (J_1^+) is correct, then v and w are identical in all extensional observational respects. Consider the classes v' and w' consisting of items exactly like those of v and w except that the distinguishing theoretical features are not truly predicable of them. Then $[v' \cong w']$ and v' is o -isomorphic with v , and w' with w . It is this *identity* which accounts for the correspondence and o -isomorphism of SENGs and PROGS. Furthermore, there are difficulties in setting up a correspondence in ways which do not seem to rely on identifications. For we decide, e.g., where physiological processes end, initially on the strength of when patients say their sensations finish, etc. Here it looks as if we have made a tentative identification. With (H_2^+) and (J_2^+) , either unsatisfactory regularity-analyses of causation are offered, in which event the hypotheses face difficulties analogous to those set out with respect to (H_1^+) and (J_1^+) , or else "cause" is characterized as a highly determinable word and central examples of causation are explained in terms of observationally related items or in terms of a story linking such items. But here we encounter a serious difficulty, namely, that without an identification like (H_6) there remain *gaps* in the observations and the story which cannot be bridged. To complete the causal story and to bridge the gaps an identity hypothesis has to be introduced. In initial formulation (H_3^+) and (J_3) suffer from shortcomings which mar most of the hypotheses, namely, obscurity. If interaction is causal interaction then (H_3^+) and (J_3^+) are confronted by the same difficulties as in (H_2^+) and (J_2^+) . It is not enough to rely on mechanical and electrical analogies like those provided by automatic control and servo-mechanisms such as thermostats and feedback circuits where the requisite explanatory details can generally be filled out. Among serious defects of (H_4^+) and (J_4^+) are that they provide no details at all of the underlying processes or of how the aspects are aspects of these processes. The analogies on which double aspect hypotheses are based, e.g., ambiguous representations, do not satisfactorily transfer to (H) and (J) cases. What is the single ambiguous item, like the drawings of the Maltese cross or duck-rabbit or the cloud-patterns, which have various aspects or interpretations? Of what are sensations appearances or aspects? In normal senses of "aspect," sensations are not aspects; so we should have to introduce a new sense of "aspect." Such a new sense can be used to advantage to collapse (J_4^+) into (J_5) . (H_6) and (J_6) have

to be carefully formulated if they are not to conflict with everyday observations. When carefully presented they are very close to (H_5) and (J_5) ; still they tend to sacrifice a main advantage of identifications: that of providing explanations, while not looking like eliminations, of familiar phenomena. Differences between various hypotheses grouped together under (J_6) , e.g., sophisticated stimulus-response and behaviorist hypotheses, only emerge under more extensive reductions of mental items. Because vague, since "over and above" can be expanded in different ways, (H_7) and (J_7) carry misleading suggestions, e.g., they suggest reductions of sense, they look like old-time analyses. If, for instance, (J_7) were replaced by "SENGs have no properties over and above those of PROGS" it would be wrong. It is correct at best when the relevant properties are restricted to extensional properties.

Accumulative reasons weigh out in favor of (H_5) and (J_5) . Therefore, as is usual, we adopt (H_5) ; and if the requisite theory P_1 were available and had some confirmation we should definitely select (J_5) .

Accepting (J_5) we can effect a *synthesis of hypotheses*. Since the intensional and significance features of SENGs or PROGS provide extra irreducible features possessed by the one but not by the other, we can combine (J_5) with a familiar double aspect hypothesis. The extra features which are specified by intensional sentence frames are irreducible because the expressions referring to each item of the contingent identity (J_5) do not have the same sense, and therefore cannot be substituted for one another in all intensional sentence frames preserving their sense. (J_5) implies this double aspect hypothesis. But the identity hypothesis is compatible with double aspect hypotheses only if the additional irreducible features are intensional or significance features.

(J_5) implies a correspondence hypothesis if "correspondence" is so redefined that v and w only correspond if they coincide but are not strictly or regularly identical. A causal hypothesis according to which PROGS cause SENGs is implied by (J_5) because (J_5) guarantees that (β) is satisfied and the requisite explanations under (β) are causal. And so on for the synthesis of other v -equivalent hypotheses with (J_5) . Each of these hypotheses, then, has picked out just some features of the contingent identity. Unless they insist (as they usually do) that the aspects they select are the only features they are all reconcilable with the contingent identity hypothesis (J_5) .

APPENDIX

Symbolize "a's statement 'p' is incorrigible" by " $I_a[p]$," and the k^a definition of " $I_a[p]$," by " $I(k)_a[p]$," "a (genuinely) believes p" by " $B_a[p]$," "a is mistaken about (the truth of) p" by " $M_a[p]$ " and "a knows that p" by " $K_a[p]$." If a's statement that p is to be incorrigible then lying, deceiving, etc., must be excluded, and a's report must be sincere; therefore we want at least to have $[I_a[p] \gg B_a[p]]$.¹⁷ Now compare definitions reached:

$I(1)_a[p]$ FOR $B_a[p] \& (B_a[p] \gg K_a[p])$

$I(2)_a[p]$ FOR $B_a[p] \& (B_a[p] = [p])$

$I(3)_a[p]$ FOR $B_a[p] \& (p \gg K_a[p])$

$I(4)_a[p]$ FOR $B_a[p] \& \sim \Diamond M_a[p]$

$I(5)_a[p]$ FOR $B_a[p] \& (B_a[p] \gg K_a[p]) \& ([p] \gg B_a[p])$

Since $[K_a[p] \gg [p]]$ and $K_a[p] \gg B_a[p]$ follow from the definition of " $K_a[p]$ " and $[B_a[p] \supset [M_a[p] = \sim p]]$ from the definition of " $M_a[p]$," it follows that $[K_a[p] \gg \sim M_a[p]]$ and $[I(4)_a[p] \supset \Box [p]]$. Thus $I_a(4)$ "disastrously" excludes synthetic incorrigible statements. $I(5)$ implies $I(1)$, $I(2)$, and $I(3)$. Since $[I(1)_a[p] = .B_a[p] \& (B_a[p] \gg K_a[p])]$ and $[I(3)_a[p] = .B_a[p] \& ([p] = K_a[p])]$ any two of $I(1)$, $I(2)$, and $I(3)$ imply the third. But $I(1)$, $I(2)$, $I(3)$, $I(5)$ are not equivalent unless knowledge is erroneously equated with true belief, i.e., unless $[K_a[p] = [p] \& B_a[p]]$. $I(2)$ and $I(3)$ are defective in that it doesn't follow from them that a knows p, i.e., we do not have for $I(2)$ and $I(3)$ $[I_a[p] \supset K_a[p]]$. Since $I(5)$ implies $I(1)$ consider $I(1)$. It has the desired consequences $[I_a[p] \supset K_a[p]]$, $[I_a[p] \supset [p]]$ and $[I_a[p] \supset \sim M_a[p]]$; but it also gives

(9): $I_a[p] \supset \sim \Diamond [B_a[p] \& [\sim p]]$.

Opposing (9) is the very plausible

(10): $\Diamond [\sim p] \supset \Diamond [B_a[p] \& [\sim p]]$; roughly, it is

always logically possible to believe what is false. If (10) is true incorrigible synthetic statements are annihilated. Using $[\Delta [B_a[p]]]$, $[\Diamond [\sim p] \supset [\Diamond B_a[p] \& \sim p]]$ follows; but to prove (10) some thesis equivalent in strength to it is required. Given (10) a weakened version of: if a can't be wrong he can't be right either, viz.

(11): $[\Diamond \sim p] \supset [\sim \Diamond [B_a[p] \& \sim p] \supset \sim \Diamond [B_a[p] \& p]]$, follows. (11) is also sufficient to eliminate incorrigible synthetic statements. For (11) is equivalent to $[[\Diamond \sim p] \supset [[B_a[p] \gg [p]] \supset [B_a[p] \gg [\sim p]]]]$ so that it follows $[[I_a[p] \& \Delta [p]] \supset [p \& \sim p]]$; therefore $[\sim \Diamond [I_a[p] \& \Delta [p]]]$. $I(5)$ is less satisfactory than $I(1)$ since $[I_a[p]]$ implies $[M_a[p] = \sim B_a[p]]$ and $[B_a[p] = \sim \Diamond M_a[p]]$ (compare (iv)). Finally these definitions of $I_a[p]$ make it plain that incorrigibility is not an extensional property.

D. M. Armstrong has proposed in comment a contextual definition of ' $I_a[p]$ ' which can be symbolized:

$B_a[p] \supset [I_a[p] \equiv .B_a[p] \gg [p]]$. But since it is equivalent, using $[I_a[p] \gg B_a[p]]$, to $[I_a[p] = .B_a[p] \& (B_a[p] \gg [p])]$ it fails to guarantee $[I_a[p] \supset K_a[p]]$. Also (9) and its consequences ensue.

Because of the consequences of (10) and (11), incorrigibilists will tend to reject both them and the arguments which support them, e.g., the Wisdom-Wittgenstein defense of (11) and the brief argument presented by Smart and others for what amounts to (10). As these defenses are commonly presented this is not a difficult task; the Smart argument, for instance, is so formulated as to rely on Hume's separability principle (Hume's separator), a principle which is valid at best under restricted conditions. (See A. Pap's criticism of the separator in *Semantics and Necessary Truth* [New Haven, 1958], pp. 190-191.) All this is not to deny that (10) and (11) do have defenses of some plausibility.

University of Sydney

and

University of New England

¹⁷ Throughout this Appendix, ' \gg ' is used for strict implication and '=' for strict equivalence.

II. MOTIVES, RATIONALES, AND RELIGIOUS BELIEFS

DIOGENES ALLEN

PART I

IN the literature about the Christian religion, there is a recurrent picture of the way religious beliefs are asserted, which may be characterized in the following way. Religious convictions are a matter of faith, but there is a certain amount of evidence for them. People may or may not come to make affirmations through a knowledge of this evidence, but there must be some evidence for religious affirmations, otherwise to assert them would be to act blindly or irrationally. In the examination of the evidence for religious affirmations, how religious people do in fact make religious affirmations is in the last analysis irrelevant because the evidence for the affirmations is the real or *ultimate* basis for them and for people's belief in them. Biographical reasons—how one comes to have faith and to remain in faith—are not the basic ground for the assertion of religious beliefs. The story of Augustine's trials in becoming a Christian as recorded in his *Confessions*, for example, is edifying, inspiring, and may even be a factor in converting some people to Christianity, but it is only Augustine's way of coming to the faith. It is not proper to cite it as a basis for Christian beliefs.

It is my conviction that biographical reasons are, on the contrary, a proper basis for the affirmation of Christian beliefs. The motives one has for one's adherence to religious beliefs are not grounds which warrant other kinds of assertions, but they are a basis for the assertion of religious beliefs. To believe on the basis of one's motives is not to act arbitrarily, blindly, or without any reason.

Moreover, as far as religious beliefs are con-

cerned, there are no reasons which are more ultimate or more basic than motives. This is the kind of basis which is appropriate to them.

I therefore argue that: (1) the response of faith to the Christian religion may be a rational and adequate basis for the affirmation of religious beliefs, and (2) there are no reasons which are more ultimate or basic than biographical ones.¹

1. *Can Motives Be Reasons for Beliefs?*

Faith is not put forward by me as a reason or ground for affirming and adhering to religious beliefs because of a consideration of the degree to which it *counts toward establishing the truth* of religious truth-claims. From the standpoint of a consideration of the degree to which the satisfactions gained in the response of faith count toward showing the truth of religious beliefs, faith is a very weak ground (if one at all) for showing the truth of a truth-claim. Nonetheless, in the response of faith a person's needs are so fulfilled that, *unless there are specific reasons which count decisively against the truth of religious truth-claims*, a person can rationally, on the ground of his response of faith and the nourishment thereby gained, adhere to these truth-claims as true. The response of faith can, without being based or undergirded by any other reason, be an adequate basis for a rational adherence to religious beliefs as true.

There are two kinds of reply to the question: Why do you adhere to Christian truth-claims? (1) I adhere to them because they awaken faith and fulfill my needs. (2) I adhere to them because there are reasons which show or count significantly

¹ This does not mean that reasoned arguments and historical evidence have no place in Christianity. On the contrary they do have a place, and in this study some of the uses of reasoned arguments and historical evidence in religion are pointed out. For example, reasoned arguments may be used to rebut a claim that a religious belief is not true or is doubtful, to remove barriers and misunderstandings which prevent one from coming to an affirmation of them, and they may be used in our attempts to conceive of God and of His relation to the universe. Historical evidence may be used to help determine what is orthodox when we are faced with rival interpretations of a Christian belief. Notwithstanding this qualification, it is still the case that religious commitment, which includes the affirmation of beliefs, is not fundamentally based on reasoned arguments nor on historical evidence.

toward showing that Christian beliefs are true. Each is of a different logical status. The latter is a reason in the sense of giving grounds for the truth of the beliefs. The former is a reason in the sense of being a person's *motive* for adhering to religious truth-claims. That one has faith and needs fulfilled by religious truth-claims does not count toward showing that they are true. It is a biographical fact about how an individual adheres to religious truth-claims.

Now it is possible that an individual's actual reason for adhering to religious truth-claims is: (1) a reason or reasons which count toward showing that the truth-claims are true, and (2) that religious truth-claims arouse faith in him and meet his needs.² On the other hand, I wish to consider the case in which an individual's actual reason for adhering to religious truth-claims may be *only* that they arouse faith in him and satisfy his needs.

I wish to show that such a motive is capable of being a basis or ground for an adherence to religious truth-claims³ without the necessity of a person relying on any reasons which count toward *establishing the truth* of religious truth-claims. In other words, faith is not *merely* a biographical fact, but it is a motive that can be an adequate basis for making religious truth-claims. I do not deny that faith is a logically different kind of basis or ground from one that consists of reasons that count toward establishing the truth of religious truth-claims. I only seek to show that faith can be a sufficient reason or ground for adhering to truth-claims, and that religious truth-claims are not in need of

another logically different kind of reason to have a basis for being adhered to.

I shall call reasons which count toward establishing the truth of religious beliefs "rationales." Such reasons can be cited by an individual as his actual reason for his adherence to religious beliefs, but they *need* not be cited by another individual who adheres to religious truth-claims. *If* faith is an individual's *actual* reason for adhering to religious truth-claims, then rationales are to him *possible* bases; ones which *some one* may rely upon for his adherence to religious truth-claims, but not *his* actual basis.⁴

2. Challenges and Rebuttals

An instance of an adherence to religious truth-claims by faith will be considered to show that motives may be a ground for affirming religious truth-claims without the need of rationales, and that such a motive does not rest on rebuttals to "challenges" to the truth of religious beliefs.⁵

Let us say that the way a person comes to believe in religious truth-claims is by exposure to the Christian community.⁶ He goes to Sunday School, he attends worship services and hears preaching, he is taught certain things about God and Jesus, he reads or hears Scripture read, he sees the way professing Christians behave, talk, and react to various circumstances, he observes and perhaps receives sacraments, he takes part in prayers. In short, he receives "training" in Christianity and he finds that he himself (either suddenly or gradually) can

² This seems to be the case with G. K. Chesterton. In a passage he considers the question: suppose that Christian doctrines contain some truth. Why not take the truths and leave the doctrines? "... Why cannot you simply allow for human weakness without believing in the Fall?" (*Orthodoxy: A Personal Philosophy* [reprint of 1st ed. of 1908; London and Glasgow, 1961], p. 141.) He gives some reasons in reply to this question and then concludes, "I have now said enough to show ... that I have in the ordinary arena of apologetics, a ground of belief. ... But I will not pretend that this curt discussion is my real reason for accepting Christianity ... I have another far more solid and central ground for submitting to it as a faith. ... And that is this: that the Christian Church in its practical relation to my soul is a living teacher, not a dead one." (P. 153.)

³ To claim that a proposition, *X*, is true, may mean: (1) there are reasons which show that *X* is true. It can also mean: (2) that one affirms and adheres to the truth-claim. To affirm and adhere to something which is a truth-claim, one is *claiming* that what is affirmed and adhered to is true. I want to show that one can rationally affirm and adhere to a religious belief *as true*, without basing one's adherence to religious beliefs on the ground that there are reasons which count toward establishing their truth.

⁴ The term "rationale" is not used with any intention of suggesting that a rationale is a "rationalization." Moreover, even if one were to be able to show that, for example, to do metaphysics is to be engaged in rationalization, as W. F. Zuurdeeg apparently claims, this would not *ipso facto* show that a particular metaphysical view or system is false, nor remove the necessity for an examination of the "evidence" for the view or system. Willem F. Zuurdeeg, *An Analytic Philosophy of Religion* (Nashville, Tenn., 1959), chap. IV, esp. pp. 137-138.

⁵ It is important to remember that the basis described is not the *only* possible kind of basis individuals may actually have for their religious commitment. This must be kept in mind, since in the following pages I speak of the motive of faith as the *actual* basis of adhering to religious truth-claims, whereas an individual may have another kind of ground for his adherence to Christian truth-claims.

⁶ This expression includes reference to a local congregation, Christians of other congregations and denominations, Christians in the past of whom one can learn, and Christian literature. The amount of exposure may vary as well.

affirm some of the things that Christians affirm. He believes in God; he can pray; his behavior is also similar to that which Christians exhibit and similar at least in some degree to the behavior of Jesus. This pattern does not seem to me to be a bizarre or an unusual one, but roughly suggests the way many people do in fact become Christians.

I wish to argue that a person could legitimately cite *only* this training and its consequences as the reason why he is a Christian: "I am a Christian because through the Christian community I have come to have faith." This would be a reason in the sense of a motive, the way he has taken which leads to his profession of faith in God. This might be the *only* reason a person has for his adherence to religious truth-claims. At the core of this motive would be the fact that he finds himself a man with faith; all the training would be the setting and medium of this awakened faith.⁷

Why would one who finds himself by this path spiritually nourished and capable of worshipping and believing in God need to give another kind of reason for his affirmation of religious beliefs? That is, a reason which is not the *actual* reason why he makes affirmations, but a reason which counts toward establishing the truth of religious beliefs and thus showing on that basis that it is justifiable for one to adhere to religious beliefs. Why is it believed necessary to give rationales for one's beliefs?

It might be said that it is necessary to cite rationales for religious beliefs because, as truth-claims, something which counts toward establishing their truth must be cited for them to be affirmed. Otherwise, a belief is held blindly or irrationally. Just because you believe is not good enough. Thus rationales must be given to show that it is justifiable to adhere to religious truth-claims, even if some individuals may believe them by faith.

My reply to this line of argument is as follows:

⁷ I do not mean to imply that "faith is without content," as if training and faith are separated. I simply wish to stress that "finding oneself with faith" is the consequence of the training.

⁸ In saying that the response of faith in this context is itself a ground of faith, I am alluding to the fact that there is a great deal of internal criticism of the practices and beliefs of Christianity by theologians, Biblical critics, and Church historians. Some of the ways this is done are described below. By internal criticism the practices and beliefs that are presented to people for adoption are carefully "screened" to prevent aberrations, superstitions, and fantasies. Thus it is possible for a belief to be ruled out and given up because (say) by historical research it is found to be "unorthodox" even though it is part of that to which one previously responded to with faith, and even if it itself may satisfy certain needs. That practices and beliefs are "screened" means that I am not saying that *anything* to which one responds with faith has a sufficient ground to be affirmed. This fact does not, however, mean that faith is itself not the reason religious beliefs are affirmed: for it is the nourishment one receives that is the actual reason one affirms beliefs and this is a sufficient reason, unless something specifically counts significantly against its truth.

⁹ Ludwig Wittgenstein, *Philosophical Investigations*, tr. G. E. M. Anscombe and R. Rhees (Oxford, 1953), p. 41. Also see Norman Malcolm, "Certainty and Empirical Statements," *Mind*, vol. 51 (1942), pp. 18-42.

A belief that has developed in the context of the Christian community is not a belief with absolutely no grounds and hence is not a blind or irrational belief. The grounds are that a man has come to have faith in response to the witness of the Christian community and in the condition of faith he finds his soul nourished. By praying, by reading the Scriptures, by fellowship with other Christians, he finds his life is beginning to conform to what Paul described as the new life. This nourishment is his ground for believing religious truth-claims.⁸ The very response of faith itself (which includes receiving nourishment) is a ground for adhering to religious beliefs as true.

Unless a specific doubt, in contrast to a logically possible one, arises to challenge the veracity of his religious affirmations, there is no necessity for further grounds for his believing them than the ground he has, viz., that he finds himself believing them. A comment of Wittgenstein's might help make this point more clear. He concludes a discussion of the circumstances in which an explanation is a final one with these words:

[It is] as though an explanation as it were hung in the air unless supported by another one. Whereas an explanation may indeed rest on another one that has been given, but none stands in need of another—unless *we* require it to prevent a misunderstanding. One might say: an explanation serves to remove or to avert a misunderstanding—one, that is, that would occur but for the explanation; not every one that I can imagine.

It might easily look as if every doubt merely *revealed* an existing gap in the foundations; so that secure understanding is only possible if we first doubt everything that *can* be doubted and then remove all these doubts.⁹

Unless there is an actual misunderstanding to be removed or to be avoided, an explanation needs no further support. In a similar manner unless there is an actual reason to doubt the truth of the beliefs

to which one is exposed in the Christian community, one's response of belief or faith is in need of no further ground.

Moreover, it is possible for there to be actual reasons for doubt (i.e., reasons which count against the truth of religious beliefs) and yet for motives alone to serve as supporting grounds. This would be justified by showing that the challenges are "ill-founded." One's grounds for belief would be that he finds himself in faith in response to the witness of the Christian community *and* that the challenges are not such as to cause him to consider his beliefs untrue. Showing "that the challenges are not such as . . ." would be a *rebuttal* but it would not be the ground that his motive of faith rests upon. Let us consider this more fully.

In the case I have specified, it has been seen that a motive can exist by itself where there are no actual reasons to challenge the truth of the beliefs to which one responds with faith. In this circumstance, it is clear that a rebuttal does not underlie or undergird a motive as a foundation. Now I wish to show that when actual doubts arise and rebuttals such as arguments, distinctions, and counter-evidence are given to rebut challenges, such may be employed, be useful, and yet not become a part of one's motivation or serve as the foundation of one's motivation. They may exist alongside motives as a different kind of reason for asserting religious beliefs. They could be cited as a reason for continuing to believe in the face of a challenge and yet not be the actual reason one believes.

Let us consider the case where challenges arise. It is possible that a person has received as part of his "training" in a Christian community some views which he later finds seriously challenged. He may have been taught, for example, any one or several of the following: a particular view of Scripture, such as that it is to be regarded as infallible on every subject with which it deals; that religion and morality are so closely related that moral living without religion is impossible; that miracles and prophecy prove the truth of the Bible; that God's existence can be proved by the very existence of the world and also by its particular arrangement; that Christianity is a world view superior to any other; that there is an entity called the soul; that there are demons. It is possible that some of these items would be mixed in as part of his motive for being a Christian, i.e., as part of that to which he reacts with faith.

The truth of all of these items has of course been questioned. A person who becomes acquainted

with a challenge or challenges could become upset, find it difficult to pray and worship, and come to believe that his religion is seriously in jeopardy. He could come to wonder whether he is able to rely upon his response of faith and he might seek a way to counter or mitigate the challenge(s) so he could believe that his faith, that he is in communion with God, is sound.

There are at least two different ways to meet challenges. One is to seek to find reasons to continue to hold to the items challenged which were taught as part of the training in a Christian community and to which he responds with faith. Another way to respond to challenges is to revise the character of the items (e.g., revise one's view of the nature of Scripture) or to consider the affirmation of some of the items as not an essential part of one's religion. These may of course be combined.

Another kind of challenge is not one that arises from doubts regarding beliefs or views learned as part of one's training, but "external" challenges such as a philosophic view which claims that religious language is meaningless. Here again one might believe that he must respond to this challenge in order to believe that his faith that he is in communion with God is not an illusion.

With both internal and external challenges, the point to be made, however it is made, is that what has been put forward does not rightly prevent one from having the belief that he has communion with God.

Now it is very tempting to say that part of the reason for affirming a religious belief or professing religion is the reason which has been used to rebut a challenge. And in one sense it is true. It is a reason for believing a religious belief or in a religion. But it is a reason in the sense of turning back a claim that would make profession of Christianity *absurd*. The actual and decisive reason for belief is still that a man finds himself believing, responding with faith to religious truth-claims. This reason has not been displaced. The reply to a challenge need not be part of the motive leading to faith. It may be cited to show that it is permissible for him to continue to believe that his response of faith is an adequate reason for faith. Someone believes because he has found himself to be a man of faith; and a rebuttal enables him to show that he may continue to believe.

Now a rebuttal, although it permits him to show that he may continue to rely on his motives, is not related to his motives as their foundation or war-

rant. It does not underlie them. To establish this point it is necessary to recall that a person who is not in the condition of doubt or confronted with a challenge does not need to have any rationale. He need not have any reasons to justify his situation as a man with faith other than that he finds himself with faith.

Shall we say, however, that this person really does need rebuttals, that he presupposes rebuttals, even if he does not know it, since challenges do exist even if he is not aware of the challenges? If we say yes, then it follows that no one ever has a religious commitment which is "really" solidly founded. For no matter how many challenges one is aware of, and however many rebuttals one has, one never is sure of having enough, since new challenges may appear. Every challenge as it materializes would reveal a gap in one's motives as a basis of affirming religious beliefs that existed all the time.

Moreover, to say rebuttals are a basis for motives suggests that there is an invisible foundation of rebuttals which emerges bit by bit as challenges arise and are rebutted; and rebuttals are the *real* ground of religious affirmations, rather than the actual way a person makes affirmations. For they establish all that is necessary to prevent it to be absurd to affirm religious truth-claims.

But challenges do not reveal gaps in motives; motives are not the sort of things which have "gaps" revealed by challenges and filled by rebuttals. In the special case we are considering, the training through which one has found that one could affirm certain things and act in certain ways is the *achievement of a goal*. One has succeeded in reaching a certain place; one has faith; one is being nourished. A challenge does not reveal that one has not attained this goal. Challenges that arise cause doubts about the truth of one's faith, but they do not reveal that the faith one has is not based upon finding oneself responding with faith and being nourished in a Christian community. Challenges do not exist as "holes" within a motive which one comes to discover; therefore rebuttals do not plug holes in a motive. *Rebuttals come and go with challenges*. Their function is to be employed to deal with challenges to the truth of religious beliefs which are affirmed by faith. They are not a permanent fixture, an invisible foundation for faith which we uncover bit by bit.

If rebuttals were an "invisible" support for a

motive, then in the case of a person who is not troubled with doubt or aware of challenges, we would have to say that such a person had no basis at all for his religious beliefs. A response of faith in the context of a Christian community would be ruled out as a ground for religious beliefs. From the perspective of a "deeper" level, we would have to say that it is not a ground at all.

Now it is not being claimed that rebuttals do not have a function and cannot be described as reasons for believing. Rebuttals function *alongside* motives. For a person who believes does not believe because there are rebuttals, but because he finds himself responding with faith. Rebuttals endorse this achievement when it is challenged; they are not the hidden but true road to the achievement.

This can be seen clearly in the case of religious people who remain in the faith even when they can find no adequate way to meet a challenge. This is a common situation, perhaps a more realistic one than the situation in which one has no doubts and is aware of no challenges. It is clear that not every believer ceases to believe when he becomes aware of a challenge. There is a time interval between a challenge being made and being rebutted. In the interval one is resting on one's motives, since a rebuttal has not yet been given. But should it be replied: Oh yes, there is another reason, viz., the belief that the challenge can be rebutted, a belief based perhaps on past successes. Suppose, however, a person cannot see any possible way to resolve the doubt that has been raised and what the challenge says seems to him to be a reason against the truth of the belief(s) he professes? If he continues to profess his belief, is he making a profession lacking a basis? Surely not. He is in the situation where he is receiving nourishment. And he has *to make the choice* whether or not this nourishment is sufficient to enable him to live with a doubt he cannot resolve or a challenge he cannot rebut. The man can grant that every challenge must be capable of being turned aside if his religion is true, but he does not see how a particular challenge can be met.¹⁰ How to live with an unresolved doubt and unmet challenge is one of the things that is taught and learned in religion.¹¹

It may appear that this distinction of motives and rebuttals and their relation is based on a rather unrealistic account. For doubts and challenges do not arise merely *after* one finds oneself with faith, but challenges to Christianity exist beforehand as

¹⁰ In fact he must believe that there is an answer to every challenge, otherwise he is not affirming truth-claims.

¹¹ One instance of an unresolved doubt and unmet challenge that we shall consider is evil. This shall be done in Part III.

well. People who take part in the life of the Church have doubts which prevent faith being awakened in them. But the distinction applies to this situation as well. Rebuttals may serve to remove doubts and challenges which act as barriers. They may serve to open the way for a person to come to have faith. But it can be seen that a rebuttal is a different kind of reason from motives by the fact that people do come to have a faith without all of the doubts or challenges being met by rebuttals. A faith can be awakened even with doubts and challenges still unmet. Thus it can be said that rebuttals do not necessarily serve as a basis or foundation for faith.

3. *Theological Screening*

It may appear that my view that faith can be the basis for adhering to religious truth-claims presumes an uncritical attitude toward beliefs and practices of the Christian religion. To say that faith is an adequate and independent basis for their assertion opens the door to any kind of fanciful conviction. Moreover, there is not a single Christian tradition, but many Christian teachings and practices, some of which are mutually exclusive. Reference must therefore be made to something else in addition to the response of faith.

My line of reply is that teachings and practices are open to examination, criticism, and revision within this scheme outlined. Christian teachings and practices can be evaluated and their adequacy supported by arguments and evidence, but adherence and confession of what a community teaches and practices would still rely on a motive, a response of faith which one finds oneself giving.

G. C. Stead's essays, "How Theologians Reason,"¹² illustrate the variety of items that enter into a theologian's consideration of the validity of Christian beliefs and in a statement of what Christianity teaches. He mentions, among other things, that one function of theology is to lay down rules to govern the original metaphors to preserve the given revelation. Another function is to give models for the proper language of worship.¹³ It sometimes uses historical investigation of the Scriptures and of the history of the Church to try to settle a dispute over a doctrine.

Besides these functions, there is a certain "open texture" about Christian beliefs, whereby theologians suggest and try out new formulations and analogies, and modify or retract previous ones.¹⁴

This brief reference to Stead illustrates the variety of the reasoning *within* the Christian religion, which indicates a careful examination and evaluation of the teachings and practices which are proposed for belief and adoption. But such reasoning is not contrary to but is quite compatible with the distinction which I have described. One is guided, for example, by the rules a theologian may suggest to determine what is properly to be said in worship. Historical study may help in determining in some cases what is and what is not orthodox. One who is in the faith will in consistency with his commitment accept results of a study which shows that something is or is not in line with the given revelation. But still there is a difference between agreeing with the results of a study and affirming what is put forward as the orthodox faith. One may affirm what has been "screened" or "scrutinized" because one responds to what it says with faith, and the "screening" is not the basis of the response.¹⁵

4. *Conclusion to Part I*

I have shown a sense in which the response of faith is a reason or ground for affirming truth-claims. For it can be a reasonable basis for the affirmation of truth-claims, as long as there are no decisive reasons which count against their truth. It is a different kind of ground or reason for making truth-claims from one that establishes their truth. Moreover, the rebuttals to reasons which count against the truth of religious truth-claims are not a support on which faith rests. They are not the *actual* basis or ground for asserting religious truth-claims, even though they "neutralize" reasons which count against their truth and which could prevent a rational affirmation of them unless they are in principle capable of being "neutralized." The *actual* reason religious truth-claims are affirmed in the case we are considering is that they awaken faith and in that condition they give a person nourishment and fulfillment. Rebuttals may be

¹² *Faith and Logic: Oxford Essays in Philosophical Theology*, ed. B. Mitchell (Boston, 1957), pp. 108-131.

¹³ *Ibid.*, p. 112.

¹⁴ *Ibid.*, pp. 115-116.

¹⁵ I do not mean to imply that the role of theology is limited to preventing fanciful convictions being asserted or to assist in the determination of what is orthodox. I have simply pointed to these functions of theology in order to reply to a charge that might be levelled against my position.

used without becoming part of the actual reason or actual basis of the affirmation of religious beliefs.

Motives, then, may be a reason for affirming and adhering to religious truth-claims. Reasons which count toward establishing their truth do not need to be given for one to assert religious truth-claims rationally. A person does not need to have such a reason or ground in order to give his adherence to Christianity or to an article of that religion rationally. He may respond directly to the claims of the gospel without logically depending on a case which seeks to establish its truth. He may by means of his motives find an adequate ground to affirm it.

Although several of my essential points have been made, there are many other essential matters which must be considered to show the truth of my thesis.¹⁶ I shall consider only two in this paper. First, it is necessary to make a distinction between those motives which may and those which may not be a reason or ground for rationally affirming truth-claims (Part II). Second, a satisfactory defense must be given for the shortcomings of theodicies which seek to rebut the challenge to the truth of religious beliefs posed by evil (Part III).

PART II

1. *Not All Motives Are Grounds*

The position that faith is a sound ground for adhering to truth-claims does not commit us to the view that *all* motives are grounds for truth-claims. I do not suppose that the satisfaction of *any* need can be a sound basis for adhering to a truth-claim. It is possible to distinguish between those motives which are not grounds and those motives which are grounds. One reason it is thought that faith is not a ground (in the sense I have argued for) is the

result of the failure to distinguish between those motives which are only sources and those motives which are sources *and* grounds for adhering to truth-claims.

John Wisdom's well-known essay, "Gods," has a passage on this matter:

I am well aware of the distinction between the question "What reasons are there for the belief that *S* is *P*?" and "What are the sources of beliefs that *S* is *P*?" There are cases where investigation of the rationality of a claim which certain persons make is done with very little inquiry into why they say what they do, into the causes of their beliefs. This is so when we have very definite ideas about what is really logically relevant to their claims and what is not. [But in some cases] . . . we have not only to ascertain what reasons there are for them but also to decide what things are reasons and how much. This latter process of sifting reasons from causes is part of the critical process for every belief, but in some spheres it has been done pretty fully already. . . . But in other spheres this remains to be done. Even in science or on the stock exchange or in ordinary life we sometimes hesitate to condemn a belief or a hunch¹⁷ merely because those who believe it cannot offer the sort of reasons we had hoped for. And now suppose Miss Gertrude Stein finds excellent the work of a new artist while we see nothing in it. We nervously recall, perhaps, how pictures by Picasso, which Miss Stein admired and others rejected, later came to be admired by many who gave attention to them, and we wonder whether the case is not now a new instance of her perspicacity and our blindness. But if, upon giving all our attention to the work in question, we still do not respond to it, and we notice that the subject matter of new pictures is perhaps birds in wild places and learn that Miss Stein is a bird watcher, then we begin to trouble ourselves less about her admiration.

It must not be forgotten that our attempt to show up misconnections in Miss Stein may have an opposite result and reveal to us connections we had missed.¹⁸

Wisdom mentions the distinction between "What

¹⁶ It is necessary to show that religious truth-claims are of such a character that the response of faith is the appropriate kind of ground for affirming religious beliefs. Otherwise, it would appear that I am only saying: we have a situation in which, as long as nothing counts decisively against a truth-claim, the satisfaction of needs is a good ground to affirm a truth-claim. I need to show that although the fulfillment of needs is not a sound ground for affirming other kinds of truth-claims, it is a sound ground for *religious* truth-claims because of their particular character. Unless this is done, adherence to religious beliefs on the basis of the fulfillment of needs becomes a matter of taste. Another closely related matter is to show that religious beliefs which make claims about the cosmos, such as the universe was created by God, are capable of being soundly based on the response of faith. Fully to make my case a description of what it is about religious beliefs that call forth a response of faith in some people is needed. I do not treat these matters here.

¹⁷ I have omitted Wisdom's footnote.

¹⁸ *Logic and Language: First Series*, ed. A. Flew (Oxford, 1951), pp. 200-201. Wisdom in this passage is defending himself against a possible objection. I am simply drawing out from the passage an objection that might be raised against me, rather than arguing against Wisdom.

reasons are there for the belief *S is P?*" and "What are the sources of beliefs that *S is P?*" This is the distinction I have described as the difference between an adherence to a truth-claim because there are reasons which count toward its truth, and an adherence to a truth-claim which is caused by matters which do *not* count toward establishing its truth. This distinction between reasons and sources suggests that because faith does not count toward establishing the truth of a truth-claim, faith is not a ground, but only a source for belief. But I believe that among sources for believing a truth-claim, we can distinguish between those motives which are *merely* sources and those which are sources *and* grounds for adhering to truth-claims (even though they do not count toward showing the truth of the truth-claims). The basis of this distinction between motives is whether or not it is the truth of a truth-claim which satisfies needs and thus motivates a person to assert (or adhere to) the truth-claim.

2. *The Bizarreness Test: Distinguishing Between Motives*

I do not deny that unworthy motives, such as an inordinate desire for attention, and irrelevant motives, such as a fear of chickens, may be involved in one's acceptance of Christianity. However, unlike the illustration concerning Gertrude Stein, we do not have just one person with a love for bird watching which might be the cause of a wrong judgment regarding a painter's work. We have many different kinds of people with many different "eccentricities" who become Christians. This variety leads me to think that unworthy motives and eccentricities, which are indeed irrelevant as grounds *in any sense* for a truth-claim, are not the only factors which enter into an acceptance of Christianity and may not even be the important ones. There may be more general causes for people responding with faith. It may be because as human beings people have needs which the Christian religion touches upon and by satisfying them awakens faith.

I intend here, however, to argue that, although the *worth* of some paintings is *irrelevant* to satisfying an interest in bird watching, *whether the belief* that there is a redeeming God *is true* or not *is relevant* to the satisfaction of human beings who feel sinful and who aspire to goodness. One can distinguish between motives for whose satisfaction the truth

of a truth-claim (*what it claims is the case*) does matter and motives for whose satisfaction this does not matter. One can often judge whether the relation between a need and its satisfaction is bizarre or proper. The *quality* of a picture does not satisfy the bird watching fancy of Gertrude Stein; that it has birds for its *subject matter* satisfies that fancy. The claim that a picture is good is not what satisfies the fancy, but that it has birds as its subject matter. Thus the relation between the bird watching fancy (the motive or source of the claim) and the claim that a picture is good is a bizarre one: it is not the claim that satisfies the need; and whether the claim is true or not does not matter to the satisfaction of the need.

The same is true of the desire for attention or a fear of chickens and an adherence to religious truth-claims. But a yearning to be righteous and the claim that there is a God who redeems is not a bizarre relationship. It is *the truth-claim* which satisfies the need (*what it claims is the case*) and whether the truth-claim is true or not matters to the person who adheres to it because of his need. This allows us to distinguish it from motives whose relation to the making of a truth-claim is bizarre.

The distinction I am drawing between motives can be seen if we take the above cases, and in each instance postulate that the person who makes the judgment is shown the source of his judgment. It would not be surprising to find a person who judged a picture to be good to give up this judgment if he were reminded that the subject matter was birds and that he had a fancy for bird watching. Likewise, it would not be surprising for one who had judged Christianity to be true to withdraw the judgment were he reminded that he had an inordinate desire to be important and that Christianity made him feel important. But if we tell one who believes in God as a redeemer that he has an inordinate desire to be cleansed a reply to this effect would not be a surprising one: "Of course, that is what I've been telling you all along; I am impure and God cleanses me." In this case the reminder of the source of the judgment does not have the tendency to cause one to consider withdrawing the judgment. In the first case one can readily see that a bird watching fancy should not be a criterion for judging the quality of a painting. The same can be said about a fear of chickens and the truth of Christianity. The desire to feel important may be considered an unworthy desire or may be praised as one of the roots of ambition. But an inordinate desire to feel important is one of the

things condemned by Christianity. It cannot be sanctioned as a basis for affirming Christian beliefs. Whereas in the last case, an inordinate desire to be cleansed neither is contrary to the Christian faith nor irrelevant to an evaluation of its truth; for one of the things Christianity teaches is the forgiveness of sins.

That religious truth-claims satisfy a yearning to be righteous, can therefore be a ground for adhering to them. For it is *because of what they claim is true* that the needs are satisfied. Were the relation between the beliefs and the needs bizarre, then the satisfaction could not be a basis for adhering to the beliefs. It would be something else which satisfied the needs, not what the beliefs themselves claimed was true.¹⁹

I am not arguing that faith is a ground in the sense of being a reason that counts toward establishing that something is true. I am defending the idea that the source for believing may be a ground for believing a truth-claim because one receives satisfaction of certain needs from what the truth-claim says is the case. The above discussion shows that the truth of religious truth-claims matters to one who responds to them because of *some* motives but not to a response caused by *other* motives. In other words, we may distinguish between those motives which can be a ground for adhering to truth-claims from those which are not. The distinction is made on the basis of whether the truth of the truth-claims matters to the needs which motivate their assertion. It is what religious truth-claims claim is true that satisfies some needs. Thus the response of faith can be one's *source* for believing and also one's *ground* for believing truth-claims. As long as the sources for believing are motives which can be satisfied only by what religious truth-claims claim is the case, they can be a *ground* for asserting truth-claims. For as we have seen, as long as there is nothing which counts decisively against their truth, they can be affirmed and asserted because they awaken faith, and nourish or fulfill needs. Not any motives which are sources for believing something will be a ground which makes it legitimate for a truth-claim to be asserted, but only ones whose relation to what the truth-claims claim is the case is proper and not bizarre.

3. Conclusion to Part II

I am not objecting to rationales—reasons which count toward establishing the truth of religious beliefs. I have tried to show that there are two kinds of grounds for adhering to religious truth-claims, and the response of faith can be a basis for affirming religious truth-claims. My intention has been to show that faith is a *kind* of ground because there is the danger that it may be thought that Christianity affirmed without a sound rationale is completely lacking in grounds of any kind.

PART III

1. Theodicies

It has been argued that the motive of faith may be a rational basis for an adherence to Christianity. But this motive can be a rational basis for the affirmation of religious beliefs as true only if there are no reasons which decisively count against their truth. Part I, however, points out that often, if not always, there are reasons to challenge the truth of Christian beliefs. The existence of evil is one of these reasons. Can it be maintained that motives may serve as a rational basis for the affirmation of Christian beliefs when there are evils, the existence of which do not seem to tally with Christian beliefs about the power, goodness, and wisdom of God? It would appear that for Christianity to be rationally affirmed, it must give an account of evil.

2. Two Different Contexts for the Defense and Evaluation of Theodicies

Many different kinds of attempts have been made to explain evil.²⁰ Christianity cannot be satisfied, however, with any account which explains evil by limiting God either in power or wisdom or goodness. Although theodicies vary in the arguments used and in the explanations given, there are at least two things which a theodicy must accomplish in order to be satisfactory. First, it must be shown that evil *as such* and the existence of God as all-powerful, all-wise, and all-good, are

¹⁹ This indicates that whether or not God exists matters to a person who affirms religious beliefs. For if the claim that *God* cleanses and redeems is not what satisfies the need for righteousness, then the relation of the truth-claim to the need is bizarre.

²⁰ For an outline of many of them, see W. D. Niven, "Good and Evil," in *Encyclopaedia of Religion and Ethics*, ed. James Hastings, vol. VI (Edinburgh, 1913), pp. 318–326.

not necessarily contradictory. Although this is a formidable task, I believe that it can be accomplished, and even persons who are committed neither to theism nor to Christianity believe that it is possible.²¹ There are, of course, those who believe that there is a logical contradiction, but this too is difficult to establish.²²

Although I do not wish to enter into detail in this matter, I do wish to mention a distinction drawn by Leibniz.²³ He says, in effect, that to avoid confusion in our reflection on evil, we must draw a distinction between what we know *a priori* and what we know *a posteriori*. On the basis of a consideration of the concept of God²⁴ we would not expect the creation to contain any evil, for the concepts of God and evil are mutually exclusive. But since we know that there is a God and since evil exists, God and evil are compatible *in fact*. He was able to draw this conclusion because of his confidence that we know that there is a God. His confidence is not widely shared today. This does not, however, destroy the value of his distinction. By pointing out that the incompatibility of God and evil is on an *a priori* level, it serves the purpose of preventing one from concluding that this incompatibility is *proof* that God and evil are incompatible *in fact*.

To show that there is not necessarily a contradiction *in fact* between the existence of evil as such and the existence of God, it can be argued that evil is in some fashion a necessary condition for the attainment of some good. An evil might be avoided, but were it avoided, some good could not be attained.

This is the line of argument pursued by Austin Farrer in discussing physical destruction and animal pain.²⁵ He argues that a material universe entails the existence of collisions and pain. You cannot have one without the other. Whether or not there must be or ought to be (and in what sense) a material universe, is another matter. But to have a material creation is to have collisions and pains,

which suggests that there may be a way in which God and evil as such are not *in fact* necessarily contradictory even though the *a priori* concepts of God and evil are mutually exclusive.

It should be noted that the purpose or ends by means of which it is justifiable to create a good thing in which evil is necessarily involved only need to be possibilities. We do not need to know what God's actual purposes are. As long as we can suggest some purposes, the mutual incompatibility of God and evil on an *a priori* level does not show that God and evil are in fact incompatible and hence that Christian religious beliefs are untrue.

The second standard which must be met by theodicies is the specification of purposes for the evils of *this world* (in contrast to evil as such). It is not generally thought necessary or possible to give an explanation of every instance or occurrence of evil.²⁶ Each class or kind of evil, as it is traditionally put, must be justified by the specification of some purpose in the attainment of which evil is involved.

Theodicies, despite the merits many of them have, all fall short of offering a complete explanation of every kind of evil and the large amount of evil that exists. An advocate of Christianity or a theism compatible with Christianity might defend their inadequacies by answering a critic as follows. If we are searching for truth and understanding, and thus are seeking an account of the world and man, then we must accept our limitations in such a search. All world views have shortcomings, so we do not need to give an account which fully explains every kind of evil and the large amount of evil in this world without some shortcomings. A theodicy can be judged to have fulfilled its function, despite shortcomings, as long as on the whole it offers the best account of evil that is available. As long as a theism which is compatible with Christianity can best explain evil, then evil does not refute such a theism nor Christianity. Moreover, we need to cope with evil, both theoretically and practically. It is a fact, and theism

²¹ John Wisdom is an example of such a person. See his article, "God and Evil," *Mind*, vol. 44 (1935), pp. 1-20. Yet he thinks it is improbable that there is a perfect God because there is so much evil.

²² Antony Flew in his essay, "Theology and Falsification," *New Essays in Philosophical Theology*, ed. A. Flew and A. MacIntyre (London, 1955), pp. 96-99, says that the contradiction is avoided by progressively qualifying and revising the meaning of statements about God until it becomes plausible to suggest that the statements about God are vacuous.

²³ G. W. Leibniz, *Theodicy: Essays on the Goodness of God, the Freedom of Man and the Origin of Evil*, ed. Austin Farrer and tr. E. M. Huggard (London, 1951), pp. 98-99.

²⁴ That is, God conceived as all-powerful, all-wise, and all-good.

²⁵ *Love Almighty and Ills Unlimited: An Essay on Providence and Evil Containing the Nathaniel Taylor Lectures for 1961* (Garden City, New York, 1961), pp. 47-49, 56-57, 75.

²⁶ This is not to deny the human need for comfort in the face of specific instances of the occurrence of evil.

or the God of revealed religion can best help us to explain evil and to deal with it practically.

This line of reply has considerable weight with one who believes that ultimate accounts or explanations of the world and man can and must be given, and who believes that such accounts *must* involve reference to a Deity who is compatible with Christianity or to God as revealed in Christianity. One way to show that they *must* involve such reference is to show that such accounts give the best account of man and the world. But it *seems* to me to be impossible to get a standard or criterion to judge what is the "best" or most adequate account that is available. For it involves the need to show in some way that there are certain "facts" which require explanations involving a perfect Deity. But it does not seem to have as yet been shown that for one to be rational one must recognize such "facts."

Thus, for one who believes for philosophical reasons that we do not *necessarily* need ultimate explanations involving reference to a Deity, the above defense of the shortcomings of theodicies offered by Christianity (or a theism compatible with Christianity) will not be accepted. For one may maintain that unless the evils of this world are fully accounted for, so that evil does not count against the views of revealed religion and a theism compatible with it, these views cannot be put forward as the best account of man and the world. A theodicy to be an adequate rebuttal to the challenge evil presents to such a theism and Christianity must fully account for evil. Evil does not decisively show that such a theism or Christianity is wrong. For it is a logical possibility that there is an explanation of evil, and the partial success of explanations that have been offered give some support to the view that there is indeed an explanation of evil as yet not fully known to us. But until we actually get a full explanation, a theism compatible with Christianity or Christianity itself cannot be put forward as the best view of the cosmos and man. Some account of the cosmos and man and evil in which there is not a perfect Deity or no Deity at all is to be preferred to one in which evil is not fully explained after the persistent effort that has been devoted to explain it and in which evil counts against the existence of a perfect Being.

There are other ways to defend the shortcomings of theodicies offered on behalf of a theism com-

patible with Christianity. A theism based on a demonstration of the existence of God, or upon an apprehension of God, or some other way of gaining knowledge that there is a God, might be able to maintain that accounts of evil need not be completely successful. For, it could say, we *know* that there is a God. Thus, evil must fit in some way, even though we do not fully know how it does. Attempts which go some way toward explaining how the evils of the world and God go together are thus of value and do not need fully to account for evil in order for us to maintain Christianity or theism.²⁷

This position, however, rests on the belief that God's existence is known. This is, of course, widely disputed by both Christians and non-Christians. Nonetheless, it is true, that *if* it is *known* that there is a God, our attempts to explain evils need not be completely successful. But without the knowledge that God exists, the shortcomings of theodicies cannot be so defended.

There is still another way to defend the shortcomings of theodicies which does not need to rely upon one's philosophic view regarding the necessity of ultimate explanations involving reference to a Deity or to the claim that we know that God exists. For theodicies may be judged according to the help they give a person in retaining a religious commitment, based on faith, to a perfect God. The standard of evaluating a theodicy is its degree of success in performing this function. A theodicy may have shortcomings and still help one maintain a religious commitment. For the affirmation of belief in a perfect God is based on the satisfaction of needs which lead to a devotion to God. It is of course necessary to show that the existence of evil as such does not count decisively against the existence of a perfect God. To maintain one's religious commitment were evil as such and God's existence shown to be *in fact* contradictory would be irrational. But it is not necessary fully to explain the evils of this world to maintain rationally a religious commitment on the basis of faith. As long as there is a good reason to believe that evil as such does not count decisively against the truth of religious beliefs, and it is admitted that there must be some fully adequate account of all the evils of this world although presently it is not known, then the nourishment or fulfillment one receives in the condition of faith allows one reasonably to maintain his religious commitment. The partial ex-

²⁷ M. C. D'Arcy argues in this fashion. *The Pain of This World and the Providence of God* (London, 1935), pp. 30-33.

planations help one to retain one's commitment and can do this despite their shortcomings.²⁸

3. *The Function of Theodicies in Maintaining a Religious Commitment*

Confronted by evil, a person may need to be fortified to retain his religious commitment. This may be accomplished by explanations. Of the two major kinds of explanation, one deals only with a single evil. For a person may be troubled by a specific evil rather than by what is usually called "the problem of evil." An example of this kind of explanation is Austin Farrer's discussion of the evils in the biological world. He points out that part of the reason some people are disturbed by the competition among animals for survival is because the behavior of animals is judged by the standard of human behavior.²⁹ Even though this, in itself, does not completely justify suffering in the animal world, it may relieve some of the horror with which it is viewed and enable a person to retain his commitment to God.

Leibniz' *Theodicy* deals with many trouble spots. For example he considers the injustice of the situation in which the wicked prosper and the just suffer adversity. He says that the gospel promises redress in an after life, and adds the interesting point that even without the gospel we would have thought of such a possibility to serve as an answer.³⁰

This additional claim introduces an interesting feature for the evaluation of theodicies in religion. It suggests that in rebutting an argument against the existence of God by the citation of some evil, it is permissible to frame *possible* ways that such an

evil can co-exist with God without showing that such a possibility is probable.³¹ For example, one of Leibniz' replies to the question, whether there exists more evil or good, is that even if there is more evil on this earth than there is good (which he thinks he has shown good reason to doubt) we still cannot conclude that this balance prevails in the entire creation. There may be so much good in the rest of the universe beyond our ken that it would reverse the balance of evil and good which prevails on earth.³²

We might be able to think of the idea of redress in an after life but such a bare possibility is a very weak reply to injustice. Also it is implausible to suggest that the rest of the universe may be so good that it can overwhelmingly turn the scales on the (*ex hypothesis*) surplus of evil on earth. The part of the universe we know best would lead us to think that it is more probable that the rest of the universe does not have an abundance of good.

"Possibilities," however, may have some value in a theodicy. For in religion, one can employ possible explanations, which have no particular backing from natural reason, if they do have some backing from religion. If one has grounds for trusting Christ, then his promise to act as judge of all men has some weight in supporting the possibility that there is redress of injustice in another life. Thus one who defends the "Mystery" can sometimes use "possibilities" to turn back charges.³³

This feature of theodicies in religion is of particular importance for the second kind of explanation which is given to help one maintain one's resolve to adhere to belief in God. The second kind of explanation is concerned with giving a purpose

²⁸ Hume apparently recognizes two contexts for the consideration of evil. *Dialogues Concerning Natural Religion*, ed. N. Kemp-Smith (Indianapolis and New York, 1947), Parts X and XI.

Hume's frequent qualification that because God's nature is considered solely as an hypothesis based on phenomena, this rules out use of possibilities or "conjectures" which would allow belief in a perfect Deity to be consistent with the evils of this world. For when God's perfection is based on an examination of phenomena, then

"these arbitrary suppositions can never be admitted, . . . Whence can any hypothesis be proved but from the apparent phenomena? To establish one hypothesis upon another is building entirely on air; . . ." (Pp. 199-200.)

Moreover, the consistency based on conjectures is effective only if one is *antecedently* convinced of the perfection of God.

"But supposing, what is the real case with regard to man, that this creature is not antecedently convinced of a supreme intelligence, benevolent, and powerful, but is left to gather such a belief from the appearances of things; this entirely alters the case, nor will he ever find any reason for such a conclusion." (P. 204.)

These are difficulties which explanations of evil encounter in the context of beliefs based on phenomena. His stress that God's nature is considered solely as an hypothesis based on phenomena suggests that the difficulties encountered in explaining evil may take on a different aspect if religion is based on faith. Hume himself does not examine, as I do, the way theodicies function within the context of faith.

²⁹ Farrer, *op. cit.*, p. 75.

³⁰ Leibniz, *op. cit.*, pp. 110-111 and p. 132.

³¹ Leibniz writes "... he who upholds the Mystery [Christianity] may answer with the instance of a bare possibility . . . without having to maintain that it is probable." *Ibid.*, pp. 118-119.

³² *Ibid.*, pp. 130-135, 287-288.

³³ I do not wish to support Leibniz' reply to the charge that more evil exists than good.

or end, the attainment of which would justify the existence of our world with the large amount of evil in it. Such a program may include among other things, a delineation of the nature and kinds of evil, proximate purposes which various evils serve, clarification of those confusions which cause an over-estimation of the extent or intensity of particular evils.³⁴ But in addition there will be an over-arching purpose for the existence of the universe which we have that justifies its creation by God.

This kind of theodicy can be exemplified by the ancient and traditional account given in the Christian Church. Evil is introduced by a disobedient angel, who, although he influences Adam and Eve, does not cause their willful disobedience. The fall of the first human creatures is transmitted (the means of transmission is described in several ways) to the rest of mankind. God in his wisdom knows not only that evil may occur in his creation, but knows that it will. It is better, however, that there be a creation with evil than no creation at all. For not only is being itself good, and free creatures able to sin better than ones not free, but God has both the resources and the will to redeem his creatures and creation from evil and confer upon fallen men a status even higher than their natural uniseful state. Such an end for his creatures is considered to be sufficient to justify a creation in which there is evil.

This of course is only an outline of the traditional view, but it is clear that it is a different kind of explanation from one which seeks to relieve specific trouble spots.

It is notorious that there are serious difficulties in this traditional theodicy and that it is open to

theological criticism on such points as the role of the devil. But this traditional theodicy may be espoused by Christianity, for a theodicy in religion need not be based on the evidence of natural reason. The account must merely suggest a possible reason why the world is as it is, and need not be actually established, nor shown to be the best account of man and the world that is available. As long as a plausible account can be suggested, the need for an explanation to enable a person to retain his devotion to God might be satisfied. Precisely how good and how plausible an explanation has to be depends in part on the person to whom it is addressed.³⁵

This stress on possible explanations is intended to make room for a certain amount of imagination or speculative guessing in a theodicy and also for a flexibility that allows for several accounts of evil. The imagination cannot be completely free, for it cannot suggest accounts which are heretical; nor can the accounts be completely fanciful or they will lack plausibility.³⁶ The use of imagination will more likely be limited to considering possibilities which are extensions of, or are accounts interpreted by and integrated with Christian beliefs. It is this relationship to Christian beliefs to which one is already committed that gives such possibilities much of their plausibility.

An example of such a use of imagination can be found in Austin Farrer's consideration of heaven and hell. He says that since the love of God revealed in Christ is so great, and since the purpose of God is to confer on men the inestimable gift of enjoying His presence for ever, the suffering for eternity of a single person is difficult to contemplate. He thus suggests that this blessing shall include the

³⁴ The fact that theodicies contain such items as these is important, not only for supporting their main argument, but also as items which may relieve specific trouble spots. For example, Leibniz' *Theodicy* is regarded as an argument that this is the best of all possible worlds. This conclusion is reached, however, in a very small space, and his theodicy is often evaluated solely on the basis of the brief argument given to support this conclusion. Although it cannot be denied that this is a feature of his work, there is another feature which is important. This feature can be described as a meeting of specific trouble spots or evils. Two such specific trouble spots have already been mentioned in the text: the prosperity of the wicked and the adversity of the good, and the balance of evil over good in the world. He considers a host of such matters in his work, generally drawing them from articles by Pierre Bayle, *Dictionnaire historique et critique*, 2 vols. (Amsterdam, 1696-97, 2me édn.; 3 vols. 1702). Leibniz' *Theodicy* thus has the feature of a series of specific trouble spots with a reply to each, loosely held together by the fact that they all deal with evil and by occasional reference to his general argument that this is the best of all possible worlds to add weight to his specific answers.

I would point to this feature of Leibniz' *Theodicy* as a reason to reconsider its merits. His work is generally not given serious consideration because it is believed that his general argument which seeks to establish the thesis that this is the best of all possible worlds is fallacious and that his thesis is absurd. This leads to a neglect of his specific replies to specific issues and the value some of them may have in relieving particular trouble spots and thereby enabling a person to maintain his commitment to God.

³⁵ This point is developed in Sect. 4 below.

³⁶ As we saw with Leibniz, to suggest that the rest of the universe may be full of so much good that it could outweigh a surplus of evil on earth, however great, is implausible.

overwhelming majority, if not all, of mankind even though the majority of mankind lives outside the range of the proclamation of the gospel and even though many are so conditioned by circumstances that they do not heed the gospel when they do hear it. He also suggests that such people will have the opportunity of responding to the love of God after this life.³⁷

Such possibilities are plausible when considered as extensions of Christian beliefs, or as hopes supported by Christian beliefs. Thus, they may serve the function of enabling a person to retain his religious commitments to God.

Another example of the use of imagination in the construction of a theodicy which may play a role in religion is G. K. Chesterton's book, *The Man Who Was Thursday*.³⁸ The story concerns an international anarchist organization and the attempt of an undercover agent, Thursday, assisted by a handful of fellow agents, to prevent the assassination of two heads of state. The author quite often crosses the line between reality and fantasy and there is a strong atmosphere of unreality even in conventional scenes. Yet, with his remarkable inventiveness, he is able to make the story as it is being read seem plausible most of the time.

The climax of the story requires the utmost effort in this regard. For the agents, who have been the hounds in pursuit of the anarchists assigned to perform the assassination, are suddenly the pursued. They are lured to a lonely place in northern France, and then, to their surprise, see in the distance a mob moving toward them whose leader they recognize as one of the leaders of the anarchist movement in Britain. They are forced to run for their lives, and in the course of their flight they are assisted by various people who shortly afterward are seen to have joined the mob pursuing them. The final blow falls when they are trapped on a jetty by the entire populace of the countryside who have gone over to the anarchists. The local police apparently desert as well for they begin to fire upon the agents. It is as if the entire world has gone mad: truth, morality, and order seem to mean nothing. It is as if chaos has conquered the earth.

The ability to make it seem even momentarily plausible to the reader that such an absurd doc-

trine as anarchy could persuade large numbers of people to join a movement requires great skill. It is crucial to the work that the reader be unable to dismiss the feeling that perhaps the world has gone mad; that perhaps chaos and irrationality have won. For a gesture of defiance, an affirmation of faith that chaos and unreason are not the truth despite their apparent victory over the earth, gains in worth and significance the more the hopelessness of the act is felt. Only if the apparent victory of chaos is made to seem real is the hopeless act of defiance and faith heightened to a pitch which is heroic.

This gesture is performed by Thursday. When he is no longer able to see a way of escape nor able to fight back, he hurls a lantern which bears the emblem of a cross into the sea out of reach of the anarchists as a final act of allegiance to reason and order.

The resolution of the story reveals that Sunday, the head of the international anarchist organization, is also the head of the secret police. That is to say, Sunday is God and has been responsible for the entire struggle between the forces of disorder and destruction and the forces of order and goodness. He rewards the six agents, who bear the six names of the week, for their loyalty in the face of the apparent victory of anarchy. He tells Thursday that he heard his declaration of loyalty in the darkest hour when all seemed lost.

The men, although they are given places of high honor and are in great comfort after their ordeal, are puzzled by this all-demanding test which Sunday had arranged for them. All but one that is. One of the men is not interested in having an explanation as long as all has come out well. He proceeds to go to sleep. Thursday professes that he can rest in comfort knowing that everything is well, but still he is curious and would like to know the explanation. But one of the men cannot even rest unless he has an explanation.

This contrast in attitudes is of some importance for specifying the adequacy of theodicies in religion. But before commenting on it, let us consider the explanation Sunday gives. One of the characters in the story unlike Sunday and the undercover agents is truly an anarchist. He really has set himself up in opposition to goodness and order; in other words, to God. The ordeal is arranged by Sunday to prove

³⁷ Farrer, *op. cit.*, pp. 95-117; and *Saving Belief: A Discussion of Essentials* (London, 1964), pp. 150-157. He balances this suggestion with an explanation of the importance of an acceptance of God's love in this life.

³⁸ *The Man Who Was Thursday: A Nightmare* (New York, 1960).

to this defiant being, Lucifer, that there indeed can be human beings who will cleave to order and goodness (to God) even when there appears to be no reward for doing so and in the most hopeless circumstances. These men are a great prize. For they show that the emergence of faithful men, worthy of keeping company with God in glory, is possible. Human beings who love God because they love order and goodness even when such allegiance seems to have the whole cosmos against it are God's desire.

Chesterton then suggests that God acts in response to the challenge of Lucifer, but it is possible to conceive of the ordeal as having nothing to do with the challenge of the devil. Instead, it could be suggested that the creation or the emergence of creatures who love God for his own sake is the purpose of the ordeal. The creation of creatures who, of their own choice and despite the evils and trials of life, are devoted to God for his own sake can be achieved only by really having them pass through evil.³⁹

I am inclined to give the story this reading, since the existence of Lucifer or a defiant one, leaves a very large gap in the explanation. Perhaps such a gap is a sound thing to have, since it suggests that the explanation of the evils of this earth takes us as far as a mystery in another realm, that of God and His relation to a non-earthly creature. But I am inclined to think that this sort of incompleteness in an explanation is less satisfactory than one which leaves the mystery with God Himself. That is, one which explains the ordeal on earth by suggesting that the emergence of faithful persons is God's desire. The mystery is that we are unable to know why God has this desire, and to judge whether this desire is justifiable, because we cannot judge whether the emergence of such persons is of such

great value that it is worth the price of the ordeal of evil in this world. Nor do we know the ultimate fate of those who fail in this ordeal. It is with the reading that would leave the mystery with God Himself that we shall be concerned.

Is Chesterton's explanation of evil in his story about an anarchist organization to be considered merely a fantasy? Does it, despite its fictional form, make a claim to be portraying something that is true or that may be true? And if it does, what supports its claim?

Chesterton's intention will not be considered. Instead, I will seek to describe a way that his book could be considered to be more than fantasy, and in what sense it can claim to deal with truth or fact.

In the context of religion, this story's explanation of evil, or the ordeals through which men pass, gets some of its plausibility from being somewhat like the actual situation in which a concern with evil arises. A concern with evil is more than a matter of intellectual puzzlement, but evils often strike us as outrageous and with a sense that it should not be so. Thursday is engaged in fighting against anarchy, coping with a secret society whose doctrines he abhors. The actual situation is one in which we are not neutral.

This is also brought out by Chesterton in his book *Orthodoxy*. He says we do not weigh the pros and cons to render a verdict on the creation. We care for life and are glad that there is a world rather than nothing. Our need, he claims, is for an account which enables us to retain this immediate love for the world, but which at the same time is able to give us the incentive to remedy its ills. To come up with an "optimistic" explanation would leave us apathetic; a "pessimistic" one would leave us in despair about righting its

³⁹ Antony Flew argues that the "free will" line in theodicies is incapable of explaining the existence of some evils on the ground that it is possible for men to be free and at the same time to be so created that they can never choose to do anything but good. "Divine Omnipotence and Human Freedom," *New Essays in Philosophical Theology* (op. cit.). Augustine in *The City of God* (*The Works of Augustine*, ed. and tr. by Marcus Dods [Edinburgh, 1871], I, pp. 450-454; II, p. 542), points out that there are several ways for creatures to be free. He says that after the revolt of some angels, those angels who remained faithful were now of such a nature that they were free but unable to sin. Likewise men in heaven will have this sort of freedom. Augustine devotes some effort to showing that this is truly a state of being free. It is this sense of freedom which Flew uses to refute the free-will defense in some theodicies. But if it is the case that God desires to have creatures with a different kind of freedom, ones who can pass through evil and still love him for himself, then Flew's argument loses much of its force. Such a love requires creatures who can sin or the ordeal is not a real test.

To argue that the existence of a creation in which there is freedom but not freedom to sin is preferable to one in which there is freedom and the ability to sin, is another matter. For it is very difficult, if not impossible, to compare such different creations. It is somewhat like comparing cows to stones or butterflies to sand. (Cf. Farrer, op. cit., pp. 56-70.) Moreover we are not in a position to know the true value of faithful souls to God to compare their worth to the trouble and evil involved in procuring them. Nor do we know God's intention for those who fail in the ordeal. That failure is indeed a terrible thing we know. How terrible and how irredeemable we do not know even though the kind of God with whom we deal gives us hope for all men.

wrongs. We need an account which will enable us to love the goodness of creation in such a way that we shall, because of that love, be able to hate it for its evils and seek to right them. We need to love the world enough to want to change it.⁴⁰

The Man Who Was Thursday has these qualities. By the explanation that some men are being tested by the apparent victory of anarchy over the world, it is suggested that evils in the world do have a place in God's economy. They are used to test and to create people who remain faithful to God for His own sake. But life on earth is not the staging of a contest between evil and goodness with nothing at stake. It matters how one bears up to evil, and how one comports oneself in the face of disorder and evil.

The Man Who Was Thursday, as a theodicy, then, gains some of its plausibility by being geared to the situation of human beings. It gives the kind of explanation which meets the needs we have as human beings living in a world with which we must come to terms by giving us a reason not to be so optimistic as to be apathetic, nor so pessimistic as to be defeated.

The strength of its claim to be seriously considered as a true account, even though it is presented as fiction, is that it is an imaginative extension of Christian beliefs which we have a basis for adhering to as true. It suggests a way that God may be regarded as sovereign⁴¹ and a role for evil which is in keeping with the Christian belief in God's love. This explanation of evil has not, however, been established as true. It is only a suggestion of the way evil might fit into God's economy, and there may be other explanations of it which are also in continuity with Christian beliefs and which are suggested by them.⁴² It need not be established as true because, in religion, one is already committed to God and has grounds in one's response of faith to religious beliefs for this commitment. A theodicy or theodicies only need to suggest a possible way to explain evils (or partially explain them) to help a person to retain his commitment in the face of evil.

4. *The Degree of Success Positively Considered*

Three standards which must be met by a theodicy in religion have already been given. First, it must be shown that the *a priori* incompatibility of God and evil is not fatal to Christianity. Second, a theodicy must lead neither to apathy nor despair. Third, an account of the evils of this world must be in keeping with Christian beliefs or be an extension of them, not only to avoid heresy, but also to give some basis for them to be considered true.

Since the concern of a theodicy in religion is with preserving personal commitment to God, its degree of success also depends upon the person to whom it is addressed. It was noted in relation to the conclusion of *The Man Who Was Thursday* that the reactions of the agents to the realization that Sunday had arranged the ordeal through which they had passed were very different. One of them was not at all interested in hearing an explanation as long as everything had come out all right. Another could rest in the realization that all was well but was still curious. Still another could not rest until he had heard a satisfactory explanation.

The point I wish to make can be made more easily if I am allowed the freedom to substitute the sentence, "All things work for good for those who love the Lord," for the comfortable situation in which the agents find themselves in Sunday's presence.⁴³

A Christian who is faced with evil may find the horror of it assuaged by the very assurance that all things work for good. He does not need to know *how* they do so. The assurance has, of course, the context of being the assurance of an apostle, and the ideas such as that God's wisdom and ways are often beyond ours, and the fulfillment of great purposes in surprising ways as recorded in Scripture reinforce this assurance.

A simple assurance may not be completely successful in the case of other Christians. Some may find comfort in it and rest their hope in it, but still find they are curious about *how* all things work for good.

⁴⁰ Chesterton, *Orthodoxy* (op. cit.), pp. 65-71.

⁴¹ One reading of Chesterton has the devil as a rival to God; but even as a rival, he is a creature. God takes this defiant creature seriously enough to show him that he is a lie by the emergence of faithful souls through an ordeal of temptation.

⁴² A Christian can use several possible explanations. For example, Chesterton's is incomplete in several ways: the relationship of the devil to God, the value of faithful souls compared to the tremendous cost of getting them, and the fate of those who fail. Other theodicies can take up these and other themes and a Christian can use one explanation to supplement another without blending them into a single account.

⁴³ They are clad in glorious apparel, having places of honor, and feel a remarkable sense of pleasure in having Sunday's company.

There are still others who *cannot* find comfort in assurances unless they are supplemented by details of *how* evil fits into God's economy. Such people find themselves unable to resist the pressure to be Christians, but they are tormented by the existence of evils. They need to have an explanation and the better the explanation is, the more the tension within them is relieved. The adequacy of a theodicy is thus in part relative to the needs of the particular Christian to whom it is addressed. But even for these people the detailed explanation need not be one which claims to be established as true. A commitment can be based on the ground of a response of faith and not upon a theodicy establishing the goodness of God.

CONCLUSION

This paper has been concerned only with the Christian religion, but its central contention that motives are grounds for truth-claims can probably be extended (*mutatis mutandis*) to any monotheistic religion, perhaps to non-monotheistic religions, and even to some nonreligious beliefs. If beliefs meet needs, pass the "bizarreness test," and have no reasons which decisively rule them out (i.e., they successfully rebut challenges), then motives can be a basis for a rational adherence to them. The evaluation of the success of a rebuttal is, however, quite complex, as we have seen in the case of the challenge of evil to Christian beliefs. For the merits of theodicies are a function of the context in which they are considered. When theodicies are evaluated in the context of adequately rebutting a challenge to a philosophic world view offered as the most adequate explanation of the universe, the standards they must meet to be an adequate rebuttal are more stringent than when they are judged for their adequacy in enabling one

to retain a religious commitment. Even when a theodicy is evaluated for its ability to rebut a challenge in the context of religious commitment, its adequacy is relative to the needs of the particular person to whom it is addressed.

Theodicies have been considered in this paper as an example of the complexity of evaluating the adequacy of a rebuttal to a challenge. This examination has, however, also resulted in showing that, by distinguishing two contexts, theodicies (such as Leibniz') may merit reconsideration for the value some of the specific replies to specific trouble spots may have even if their general argument is wrong.

To offer motives as grounds for beliefs (whether Christian, non-Christian, or otherwise) it would, however, be necessary to show in each case that motives are the appropriate kind of ground for their assertion by showing that reasons which count toward establishing their truth are rationales.

The distinction between motives and rationales should suggest why people who continue to hold beliefs when their grounds (which apparently established or significantly counted toward establishing their truth) are destroyed, rightly believe that they are not acting irrationally. For when proofs for God's existence, for example, are shown to be invalid, it is only a rationale that has been destroyed, not the actual ground for adherence to belief in God.

It may be asked: "How does one choose between rival religious beliefs on the basis of this account of motives as a rational ground for beliefs?" On the assumption that rival beliefs are both shown to be beliefs which can appropriately be based on motives, pass the "bizarreness test," and adequately rebut challenges, then choice between rival beliefs depends on which religious beliefs (and practices) meet an individual's needs.

III. LYING

FREDERICK A. SIEGLER*

TYPICALLY, when a man lies he says what he knows to be false in an attempt to deceive the listener. I shall single out six features embodied in the typical case of lying and investigate the relationship between each feature and the correct application of the concept of lying. One might be inclined mistakenly to believe that what is embodied in the typical case of lying represents necessary conditions for any and all cases of lying, but I shall try to show some connections between each of the features and various locutions which make up the cluster surrounding the concept of lying. The six features are that the liar must: (1) say something, (2) intend to deceive, (3) say something which is false, (4) say something which he knows to be false, (5) believe that what he says is false, (6) communicate. The locutions at issue are as follows:

1. *A* was lying.
2. *A* was lying to *B*.
3. *A* was telling a lie.
4. *A* was telling a lie to *B*.
5. *A* lied.
6. *A* lied to *B*.
7. *A* told a lie.
8. *A* told a lie to *B*.
9. What *A* told was a lie.
10. What *A* told to *B* was a lie.

1. *Saying something.* Surely if *A* has written in a letter that he is ill, it is correct to describe what he has done as saying in a letter, saying in writing, etc. And again, a mute who uses sign language correctly can be said to be saying something with his hands, or even speaking with his hands, or talking in sign language. Again, in sending a message in Morse code or semaphor, *A* can be said to have said something. Speaking or saying

something seems to be connected with conveying meaning through conventional signs not simply exclusively with uttering the sounds of a spoken language. For the sounds of a spoken language are simply one of several conveyors of meaning in a language. If this is so, then although uttering words is not a necessary condition for lying or telling a lie, it may be that saying something or doing something which can be described as saying something is a necessary condition for lying or telling a lie. This way of putting the matter will allow us both to speak of saying something as a necessary condition for lying or telling a lie, and to distinguish cases which come close to lying or telling a lie from genuine cases. For suppose that *A* pretends to have a limp hoping to deceive *B* into believing that he is incapacitated. *A* may know that he is not incapacitated, intend to deceive, but it seems that unless *A* does something which can be described as saying something, he cannot be said to be lying or telling a lie.

Now it may appear that if *A* says something and *B* hears it, and *A* intends to deceive *B* into believing something is false, then *A* necessarily has lied, but that is not so. Suppose that *A* and *B*, in planning to murder Jones, arrange for *A* to give *B* the order to shoot when Jones is at the appointed target spot. Suppose that *A*, to deceive *B* and thereby to get Smith shot, gives the order "shoot" when he sees Smith at the target spot. *A*, in giving the order, has said something, but it does not seem correct to say that *A* has lied to *B*. He tried to deceive him by giving the order at the wrong time, but he has not lied.¹ Suppose that the arrangement was for *A* to signal, not to order, when Jones is at the appointed spot. The signal is "Bingo." Even here, if *A* gives the signal when he sees Smith, it is not clear that he has, in signalling, lied to *B*. He

* An early version of this paper was read at the Western Division of the American Philosophical Association, May, 1964.

¹ It might be argued that in virtue of the context, *A* could be described as saying "Shoot now! Jones is here." Cf. *infra*, section 5. It might also be argued that saying something or doing something which can be described as saying something is not a necessary condition for lying. Both Augustine and Aquinas seem to speak of lying and lies (they do not distinguish as I shall) when there is no speaking or description of action as saying something. Aquinas agrees with Augustine's statement, "To pretend is not always a lie: but only when the pretense has no signification, then it is a lie." (Augustine, *Liber II De Quaestione Evangelii*, qu. 51, in *princ.*) Aquinas and Augustine defend Christ against lying when he "made as though he would go farther" (Luke: xxiv) on the ground that he was signifying something figuratively, namely, "He was about to go farther away from

gave the signal at the wrong time to deceive *B*, but in signalling he was not lying, nor telling a lie, and what he told was not a lie. This might suggest that an important, if not a necessary, condition for lying is saying something *which is false*, or perhaps saying what one *believes to be false*.

2. *Intending to deceive*. Suppose that *A* is forced by *C*'s threat to tell *p* to *B*. *A* knows that *p* is false and that *C* intends that *B* be deceived into believing it. *A* hopes that *B* does not believe *p*, but he tells him and does nothing to prevent his believing *p*. Can we say that although (or because) *A* was forced to lie or to tell a lie to *B*, *A* did not intend to deceive *B*? If we can say this, then, intending to deceive is not a necessary condition for any of the locutions. But intention to deceive does have a connection with this case. At least it is true that what he said was intended to deceive *B*, even though it was not *his* intention, for surely *C* intended that what *A* said, namely *p*, deceive *B*. This suggests that if any of the locutions apply there must be intention to deceive, although perhaps not necessarily the intention of the subject of the locution.

In this case, then, all of the locutions apply, and *A* did not intend to deceive *B*. But is there an element of intention ascribable to *A*? Suppose *A* admits that what he said was intended to deceive *B*, but that it was not *his* intention to deceive *B*. Exactly what is the force of his disclaimer?

A might hit *B*, knowing what he is doing, and unintentionally kill him. But he could not knowingly hit *B* and unintentionally touch him, for hitting involves touching and if *A* knows that he is hitting he knows that he is touching, and consequently his touching cannot be unintentional while his hitting was intentional. Not quite similarly *A* might promise to give *B* a dollar and unintentionally offend him, but he could not promise to give him a dollar and unintentionally offer assurance that he would give him the dollar, for to promise is, *inter alia*, to offer assurance that one will do something. *A* might hope that *B* does not take his promise seriously, but that is close to hoping that *B* does not take his offer of assurance seriously, because in promising, *A* does offer assurance. He cannot claim that he did not intend to

offer assurance; and if *B* is assured, then since *A* knows this is the most likely consequence of his offering assurance, he cannot claim that his assuring *B* is unintentional or simply an unintentional consequence of his having promised (offered assurance).

Now, in the case of lying which we have considered, it does not seem right to say that *A* intended to deceive *B* because that would suggest, falsely, that deceiving *B* was his aim or goal in speaking. Yet it seems not quite right to say that *A* did not intend to deceive *B*, for that would suggest, falsely, that he did not realize that what he said might well deceive *B*. Again if *B* was deceived by what *A* said into believing *p*, then it seems wrong to say that the deceit was unintentional, and perhaps somewhat misleading to say that it was intentional. But because *A* knew that he was lying, telling a lie, and that what he was saying was a lie, and consequently he knew that it would likely deceive *B*, we might say that if *B* was deceived, *A*'s deceiving *B* was an intentional action. This is to emphasize that *A* knew what he was doing, namely, speaking falsely to *B* knowing it was likely to deceive him.

The case would be different if *A* not only hoped that *B* would not believe *p* but did something to prevent him from believing *p*. For example, suppose that *A* winked as he uttered *p*. The winking is done not only with the hope but with the intention of preventing *B* from believing *p*. Consequently, if *A* winks as he says *p* it is not entirely clear that *A* has lied to *B* (1, 2, 5, 6), since surely the winking is part of what *A* was doing, and what he was doing was not with the intention of deceiving *B*, and was not intended to deceive *B*, and if he did deceive *B* it would have been unintentional, since *A* intended to keep *B* from being deceived. But on the other hand, we might distinguish between what he *said* and what he *did*. What he said was false, and it was intended (by the threateners) to deceive *B*, and *A* knew this, although what he did, namely speak and wink to warn *B* against believing what he said, was intended (by *A*) to prevent deception. Consequently we might say that what *A* told was a lie (3, 4, 7, 8, 9,

them by ascending into heaven, He was, so to speak, held back on earth by their hospitality." (Aquinas, *Summa Theologiae*, 2a-2ae, Q. 111, art. 1 ad 1; also Augustine, *Contra Mendacium ad Consentium*, 28.) I have argued that where there is nothing said nor describable as something said, a man can attempt to deceive but he cannot be said to lie or tell a lie. For the most part Augustine seems to agree with this in *Enchiridion*, 18, where he says that the very essence of lying is having one thought in one's heart and another on one's lips. Cf. *Contra Mendacium*, 12, "a lie is a false signification by words"; and also *Contra Mendacium*, 10, "a false statement uttered with intent to deceive is a manifest lie." Cf. Aquinas, *ST* 2a-2ae, Q. 111, art. 1, "The essential notion of a lie is taken from a formal falsehood, from the fact, namely that a person intends to say what is false."

10) even though *A* did not lie (1, 2, 5, 6). This would suggest, I think, correctly, a close connection between a *lie* and something *said*, and a close connection between *lying* and something *done*. This is not to suggest that saying something is not doing something. Of course saying something is doing something. It is to suggest that the noun which labels what was said (a *lie*) may be applied when the correlate verb which labels what was done (*lie*) ought not to be applied. If *A* winked as he told *p* to *B*, he would not think that *B* would likely believe *p*. And consequently there is less plausibility in saying that if *B* did not notice or take seriously the wink, *A*'s deceiving *B* was an intentional action. On the contrary, it was contrary to what was intended by *A*. I think that this shows that *A* could tell a lie to *B* when his act of deceiving *B*, if *B* is deceived, is *not* intentional. For consider the following case: Suppose *C* pays *A* to tell *p* to *B*. Both *A* and *C* know that *p* is false. *C* wants *B* to believe that it is true. *C* is therefore paying *A* to lie or tell a lie to *B*. Now it seems to me that *A* could tell *the* lie, namely *p*, to *B* without lying to *B*. Suppose *A* arranges for *D* to tell *B* that *A* is going to tell him (*B*) a lie at 2.00 P.M. Tuesday, and at that time *A* tells *p* to *B*. It seems correct to say that *A* told the lie to *B*, but it does not seem quite right to say that *A* lied to *B*.

Again, suppose that *A* did not know or believe that *p* was false and did not know that *C*, who paid *A* to tell *p* to *B*, knew it was false and intended to deceive *B* through *A*'s telling him *p*. I shall later discuss whether *A* must know or believe that what he says is false if he is lying or telling a lie, but here I am only concerned with whether *A* could lie or tell a lie when he does not intend to deceive *B*, and when, if *B* is deceived, *A*'s action of deceiving *B* is not an intentional action. Here we could say that *A* passed on *C*'s lie, and perhaps told *C*'s lie, or told the lie to *B*, and surely neither did *A* intend to deceive *B* nor could *B*'s being deceived be ascribed as an intentional action of *A*'s. This again suggests a connection between something's being a lie and its being false, and a connection between somebody's lying and his intending to deceive (*his* being false).

The conclusions derived from examination of these cases seem to be as follows: If *A* lied or told a lie (all the locutions), then at least somebody

intended that *B* be deceived into believing *p*. To that extent, intention to deceive is necessarily involved in lying or a lie. If *A* lied (1, 2, 5, 6) then, at least if *B* is deceived, *A*'s action, namely, deceiving *B*, is an intentional action. And *A* can pass on or tell a lie, or what he told can be a lie (3, 4, 7, 8, 9, 10), when *A* did not intend to deceive, and if *B* was deceived, then *A*'s action of deceiving *B* was not intentional. In this last case, e.g., when *A* winked, we distinguished between what *A* did, which was more than utter certain words, and what *A* uttered. (It might be thought that *A* cannot unintentionally deceive *B* but only unintentionally mislead *B*, but on the contrary, suppose that someone issued certain utterances and later lamented that he had forgotten to precede the utterances with negation signs. Imagine God lamenting such an omission about the ten commandments.)²

3. *What he says must be false.* Consider the case of Pablo Ibbieta in Sartre's story, *The Wall*. Pablo was asked where Ramon Gris was hiding and Pablo in an attempt to deceive the authorities told them that Gris was hiding in the cemetery. He was quite certain that Gris was hiding somewhere else. He told them what he believed to be false in an attempt to deceive them. But did he thereby lie? It is not clear, for one element of the typical case is missing, namely, the falsity of what is said. Ramon Gris, by chance, and completely unknown to Pablo, was hiding in the cemetery. Surely, it is tempting to hold that it is not a necessary condition for lying that what he said be false, since in this case he did believe that it was false, and said it with the intention of deceiving the authorities.³ Such behavior would be excellent grounds for mistrusting his honesty, and if he did this often it would be excellent grounds for calling him a liar. For it is just that sort of thing that liars do. And after all if a man did this regularly we would not want to exonerate him from being a liar simply because he is also a fool. On the other hand, it does not seem quite right to say that in saying that Gris was in the cemetery, Pablo was telling or told a lie although what he told was in fact true or the truth. Again, it is queer to say that although what Pablo told the authorities was a lie, what he in fact told them was also the truth or true.

If Pablo were asked immediately afterward

² Augustine, *Enchiridion*, 18, 19, 20; and Aquinas, *ST*, 2a-2ae, Q. 110, art. 1. Both require intention to deceive for lying, and something's being a lie. But neither distinguishes between lying and something's being a lie.

³ Cf. Augustine, *De Mendacio*, ch. 3: "a person is to be judged as lying or not lying according to the intention of his own mind, not according to the truth or falsity of the matter itself."

whether he had told a lie, and he were to answer honestly, he would say that he had. But then when he found out that what he said was in fact true how would he refer to what he said in retrospect? Could he say "I told them a lie, but it turned out to be true"? or "I thought that I had told (was telling) them a lie but it turned out to be the truth"? Did he mistakenly think that he had told a lie? Is it false that he told a lie? It sounds better to say that he *could not have been* telling a lie, than to say that he simply *did not* tell a lie, for the latter suggests, falsely, that he told them *the truth* (as opposed to "he told them what was, in fact, and contrary to his belief, true").

A solution for the Pablo problem will involve making use of the distinction between lying and telling a lie. It seems sensible to resist saying that Pablo told a lie, told a lie to the authorities, that what he said or told was a lie, and this resistance seems to be due to the closeness between the falsity of what was told or said and telling a lie. Consequently it may be that a necessary condition for correctly saying that *A* told *B* a lie, *A* told a lie to *B*, what *A* told was a lie, etc., is the falsity of what *A* said. But even if this is so, it does not follow that the falsity of what *A* said is a necessary condition for correctly saying that *A* has lied, was lying, was lying to *B*, etc. And if that is so, it is because of the closeness between *A*'s saying what he believed to be false in an attempt to deceive, and *A*'s lying.

There are cases in which this difficulty about the falsity of what *A* says does not arise. Suppose *A* is asked whether he believes that Johnson will be re-elected in 1968, and he answers that he does believe this. In saying "I believe that Johnson will win in 1968" *A* is lying if he does not believe this. Here there is no distinction between the question "Does *A* believe that Johnson will win in 1968?" and "Does *A* believe that what he said was true?" For *A* cannot believe that Johnson will win in 1968 and not believe that what he said was true. So in these cases the question of whether the falsity of what *A* says is a necessary condition for lying reduces to the question of whether *A* must believe that what he said is false, if he is lying. And we have not yet turned to that question.

Consider now an attempt to defend the view that the falsity of what *A* says is a necessary condition for lying. It might be held that whenever *A* lies there is some correct description of what he said which would include the feature "What he

said was false." If Pablo said that Gris was in the cemetery, can he be described as saying that he believed, thought, had no doubts that, Gris was in the cemetery? He did not *say* any of these things. He can be described as *expressing* a belief that Gris was in the cemetery. If we could describe Pablo as not only expressing that belief but also ascribing that belief to himself, expressing *his* belief, i.e., saying that he held that belief—then we should have a description of what he said such that what he said was false. And if this is correct then we should have no grounds for saying that the falsity of what he said is not a necessary condition for lying, since this was the crucial case which was to show that *A* could be lying when what he said was not false.

But this analysis is not correct. Pablo said that Gris was in the cemetery. On the basis of his saying this, the authorities could infer that he believed this, but this is not because Pablo *said* that he believed it. He did not ascribe that belief to himself, nor did he say that he held that belief. We might say that he *expressed* that belief, but that is not the same as his *saying* that he believed it. He might express a belief when he does *not* so believe, in which case he does not express *his* belief.⁴ A man can express a belief by saying "I believe that *x*" or by saying "*x* is so." But saying "*x* is so" is not saying "I believe that *x* is so."

Consequently it is simply false that what Pablo said is false. Nevertheless, it is not simply false that there is *something false* in what Pablo said. Could we say that Pablo expressed a false belief? That would be misleading at best since it suggests what is false, namely, that what Pablo said was false. We could say that Pablo falsely expressed the belief that Gris was in the cemetery, for that does not suggest the falsity of *what* he believed, namely, what he said; but rather it suggests that he was not honest or truthful in expressing this belief. This seems correct, and it comes to the same point we considered earlier, that the falsity of what is said is perhaps a necessary condition for telling a lie but not for lying, and that lying requires believing that what was said is false, or falsely expressing a belief.

It does seem to be the case that a necessary condition for lying is either saying what is false or implying something that is false. For, in the case of Pablo, in saying that Gris is in the cemetery, he does imply that he believes that, and therefore

⁴ Compare: In saying "I doubt (am sorry, apologize)," he expressed doubt (sorrow, apology) but he was insincere. He does not doubt (is not sorry, feels no sorrow, is not apologetic, does not feel apologetic).

what he implies is in this case false. And it seems that in every case of lying this must be so.

Now I should say that cases like that of Pablo are interesting because of their rarity and the consequent conceptual problems which they bring. It seems clear that the concept of lying or telling a lie does not very happily apply in such a case, and when such a case arises it is likely that these concepts simply would not be used in describing the case. We should say rather that Pablo attempted to deceive the authorities by telling them what he mistakenly believed to be false. Surely that describes the case clearly and correctly without the use of any notion of lying or telling a lie. Consequently, in one way, we should conclude that the falsity of what is said is a necessary condition for lying or telling a lie. Yet if we ask specifically whether it is true to say that Pablo lied or told a lie it seems uninteresting simply to say "no" on the ground that what he said happened to be true. Where, for example, something hangs on the application of the concept of lying, as in a court of law, it seems that somewhat more subtle discriminations and analysis are required. In fact, the law provides just this additional analysis, for if *A* testifies that something is true when he believes it to be false, then even though what he says is true, in English and in some American law, he may be held for perjury. Consider, in this regard the following case. *A*, the sole witness of a murder, has no doubt that *B* killed *X*, but because of personal hatred of *C* testifies that *C* committed the murder. Suppose that *C* has no alibi and is convicted and sentenced. Then suppose that *A* confesses to having lied in his testimony, and he is tried and convicted and sentenced for perjury. But then suppose that *C* confesses the crime and reveals conclusive evidence that he did commit the crime. Does it seem just to retry and exonerate *A* because his testimony was true although not truthful? I should think not. Consequently, although the case of Pablo is not a clear case of

lying, if examined, it seems more plausible to subsume it under the rubric of lying but not of telling a lie. He bore false witness but spoke what was true. This suggests the apparently paradoxical conclusion that if a man is bent on evil he may find it easier to become a liar than a teller of lies.⁵

4. *He must know that what he said is false.* Clearly from what has been said about Pablo it is not necessary that he knew that what he said is false in order for him to be lying, for it was not false and so he could not have known it to be false. And if it was false, then what he said was a lie, and he told a lie even if he did not know it to be false but simply believed it to be false. Consequently this condition is not necessary for lying, etc., or telling a lie, etc.

5. *He must believe that what he said is false.* Suppose that *A* is hired to broadcast for a radio station which is, unknown to *A*, engaged in propaganda activities. Suppose that much of what *A* is told to say is false, although *A* either does not think about the truth value of what he says or he believes that what he is saying is true. Now it seems true to say that although *A* does not realize it, many of the things he says are lies. In this case believing that what he says is false does not appear to be a necessary condition for correctly saying that what he says is a lie. Could we not also say that he has been telling lies to the public? If so, and I am inclined to think it is so, that is because of the close connection between the falsity of what *A* says and the notion of telling a lie. Of course, it is telling lies as opposed to simply giving incorrect information when he does not believe that what he says is false, in this case, because of other circumstances peculiar to this case. In particular somebody, the writers for the program, write what they believe to be false, and with the intention of ultimately deceiving the public. And again, *A* could deny that he was lying, since he did not believe that what he was saying was false. He

⁵ Augustine and Aquinas agree that the falsity of what is said is not a necessary condition for lying, but since they do not distinguish between lying and telling a lie (or something's being a lie), they also think that falsity is not necessary for telling a lie or something's being a lie. Cf. Aquinas, *ST*, 2a-2ae, Q. 110, art. 1: "If . . . one utters a falsehood formally, through having the will to deceive, even if what one says be true, yet inasmuch as this is a voluntary and moral act, it contains falseness essentially and truth accidentally, and attains the specific nature of a lie." But Augustine seems to provide the rudiment of the distinction I have suggested between lying and telling a lie (or something's being a lie). For in *Enchiridion*, 18, he says, "the man who says what is true, believing it to be false, is, so far as his own consciousness is concerned, a liar. For in saying what he does not believe, he says what to his own conscience is false, even though it should in fact be true; nor is the man, in any sense, free from lying who with his mouth speaks the truth without knowing it, but in his heart wills to tell a lie. . . . Though his statements may be true in fact (he) has one thought in his heart and another on his lips; and that is the very essence of lying." Augustine has not said that in such a case the man told a lie, or that what he said was a lie. And when he does speak of something's being a lie, e.g., in *On Psalm 5*, 7; he says, "a lie is contrary to truth." And later, "to speak of what is not, is to tell a lie." Though Augustine does not explicitly make the distinction, he seems naturally inclined to make use of it in his analysis of the concept of a lie.

could say "I was not lying; I believe all that stuff." But, in the case in which *A* says something which is not itself false but implies something which is false, it must also be the case that *A* believes (in this case falsely) that what he says is false. That seems necessary for saying that Pablo was lying. To test this condition, consider a case in which *A* does not believe that what he says is false.

Suppose that *B* asks *A* whether he owns a television set, and suppose also the context clearly shows that *B* would like to watch a program on somebody's set. *A* says "No, I do not own one." And in fact that is true but there is a television set at his home which was left for *A*'s use (or *A*'s roommate owns one which *A* uses). In saying what he did, *A* implied that there was no television set at his disposal and that is false. And *A* said what he did to deceive *B* into believing that he had no television set at his disposal. Here what *A* said was not false and he did not believe that what he said was false. If what was said under Section 3 is true, then *A* was not telling a lie. But did *A* lie? I think that we should not say that *A* lied to *B* just because, unlike the case of Pablo, *A* did not believe that what he said was false. He was being deceitful in implying what he believed (knew) to be false, but he was not lying.

But now suppose *B* is waiting for her boy friend, *C*, who owns a green Dodge convertible with white wall tires, to come for a visit. *A* sees someone she knows not to be *C* arrive in front of the house in a green Dodge convertible with white wall tires, and says to *B*, intending to mislead *B* into believing that *C* has arrived: "A green Dodge convertible with white wall tires has arrived." This case seems to be similar to the former one, in that *A* did not say what she believed to be false but implied what she believed to be false, and consequently we should not say that *A* had lied to *B*.

But now suppose that *B* is waiting for her boy friend, *C*, to arrive, and *A*, who knows that *C* has taken a liking to the lady who lives next door, sees *C* approaching the next door. *A*, intending to mislead *B* into believing that *C* has come for her, says, "*C* is at the door." Did *A* lie or tell a lie to *B*? Did *A* say something which is false and/or say something which she believed to be false? (The question is whether *A* said or implied what is and/or what she believed to be false.) Suppose that *A* denied that she lied or told a lie on the grounds that in saying what she did she was referring to the next door, and thereby she told the truth. Is such

a denial open to *A*? Or should we say that she cannot by an act of intention make the referent of "the door" the door next door and not their door?

Consider a case which is more blatant, and which, if analogous in the relevant respects, will help settle this one. *B* is waiting for her boyfriend, *C*, to come, and *A*, seeing the landlady at the door, says, "Your boyfriend is here." She says this in an attempt to get *B* to believe that her boyfriend has arrived. How could *A* deny the accusation that she lied on the grounds that by "your boyfriend" she meant to refer to the landlady? Clearly *A* does not have control over what is meant by "your boyfriend." But more important, even if she did have such control, in this case it would be false to say that by "your boyfriend" *A* is referring to the landlady. On the contrary *A* is referring to *C*, the boyfriend. But this is perhaps not the right way to put things. There is, of course, some confusion about the notion of a referent. Clearly in saying "your boyfriend," *A* intended that *B* understand her to mean *C*, and moreover, that is what "your boyfriend" *does* mean in that context. And accordingly, what *A* said is false and she knows it. Consequently, *A* cannot in honesty deny that that is what "your boyfriend" meant. And even if we do allow (I think, wrongly) for *A*'s insistence that *she* meant the landlady, then what she meant is not what she said. What she said means that *C* is here, and *A* knows that is false. Consequently, in saying that, *A* lied and told a lie.

Is this case similar in relevant ways to the previous one? Should we not say that what *A* said *meant* that *C* is at their door, and not that *C* or somebody like *C* is at some door or other? And again, even if we allow that *A* might say that *she* meant "the next door" still what she *said* meant that *C* was at their door. That was false and *A* knew it. Consequently, *A* was lying and told a lie. In the television and green convertible cases, what *A* said implied what was false and what *A* believed to be false. But in the door and the landlady cases what *A* said was false and *A* believed that it was false.

Strawson⁶ once defended the view that a sentence whose subject term lacks reference is neither true nor false, and although he has amended the view in certain ways it seems to be false and not so clearly amendable as Strawson suggests.

If a beggar stops me on the street and asks for some money by holding out his hand and saying, "All my children need medical attention," it seems

⁶ *Introduction to Logical Theory* (London, 1952), pp. 173-178.

highly suspect to say that when I find out that he has no children I cannot correctly say that he told me a lie. And there is something suspect in suggesting that the beggar could retort, "I did not say that I had children, I simply said that all my children need medical attention."

Consider the following retort: "Suppose you were asked in a court of law: 'Did he actually say that he had children or did he simply say that all of his children were ill?' Clearly he said only the latter. Consequently what he said is neither true nor false. He may have implied or suggested that he had children but he simply did not say this."

But how far does this objection travel? Suppose *A* says "My wife is ill." Did he actually say that he was married? Did he actually say that he had a wife? Surely he did not simply imply that he had a wife. (1) Suppose *A* says, "I have an ill wife." He did not say that he was married. He did not even say that he had a wife. But this comes to no more than that he did not actually use certain words. Suppose *A* said, "Ma femme est malade," and in an English court you were asked, "Did he actually say, 'My wife is ill?'"

(2) Compare: *A* and *B* are on a sinking ship. *A* cannot swim and *B* can. *A* says to *B*, "Save me and I will give you every penny that I have." *B* saves *A* and then *A* gives *B* three pennies which are all the one-cent coins he has. Did *A* say that he would give *B* all of the money that he had? Did he imply that he would give *B* all the money he had?

(3) Suppose *A* said, "Save me and I will give you all the money I have." And then after being saved he reveals his empty pockets and bare check-book. *A* is bankrupt. Did *A* fulfill his promise? Did he imply that he had money? Did he suggest it?

In (3) it seems clear that *A* did suggest or imply that he had money to give, and because of this he was trying to deceive or mislead *B*. And in a straightforward sense he did fulfill his promise. But of course fulfilling his promise does not excuse him or eliminate the description, "He tried to deceive (or did deceive) *B*." And if we have to decide between assigning a truth value or neutrality to what *A* said, then *A* made what he said true. Or if this is a bit odd sounding, put it this way: After the procedure it would be true rather than neutral to say, "*A* gave *B* all the money he owned."

But in (2) it seems right to say that *A* not only

¹ *Ibid.*, p. 178.

tried to (or did) deceive *B*, but perhaps also that *A* told a lie to *B*, for he said he would give him every penny he had, and that means (usually, and it *did* mean in this case) all the money he owned.

Now in (1) when *A* said "I have an ill wife" he did not say, "I am married" in the sense that he did not use those words. That is to say, he did not utter those words. But that is not sufficient for denying that he said that he was married, such that what he said was false. Similarly with "I have a wife."

Finally, in the first case, most crucial for Strawson's thesis, *A* did not utter the words "I have children," and the lawyer's question seems to come to little more than this issue. It is not quite clear here as it seems to be in the other cases, especially in (3), but it does seem wrong to say that *A* implied, suggested, hinted, indicated, insinuated, that he had children. And it seems more nearly correct to say that he *said* it, but not quite that he affirmed, declared, or even stated it. Perhaps here it is clearest to say that *A*, as it were, said that he had children. This locution is slightly suspect for its obscurity, but it seems appropriate in that it does allow for what seems to be a natural feeling that in saying what he said, *A* was telling a lie, and that therefore what he said is not neutral with respect to truth or falsity. And this is to say that what he said is not neutral simply with regard to honesty or dishonesty, but more than this, for *A* was not simply being dishonest in saying what he said. But if one has to decide whether with respect to truth or falsity as opposed to neutrality of what he actually said, what he said was false.

Strawson considers a case which comes close to telling a lie and he seems to think that it covers the cases I have considered above. "[If] someone, with a solemn face, says 'there is not a single foreign book in his room' and then later reveals that there are no books in the room at all, we have the sense, not of having been lied to, but of having been made the victim of a sort of linguistic outrage. Of course he did not say there were any books in the room, so he has not said anything false. Yet what he said gave us the right to assume that there were, so he has misled us. For what he said to be true (or false) it is necessary (though not sufficient) that there should be books in the room."⁷

But if the man did not immediately reveal the truth, and said what he did to deceive, it is not entirely clear that we should think of what he said as a linguistic outrage as opposed to a lie. And surely

Strawson's account of this case will not do for the cases we considered above. Strawson has acknowledged this in a later article⁸ and there he claims that his analysis is an account of the *primary* uses of the terms "true," "false," etc.

What exactly is meant by the distinction between "primary" and "secondary" uses of terms I have never quite understood. Surely, Strawson's account applies to many and, perhaps, we should say, to typical cases of appraising speech acts. This is supported by the fact that in the lie cases we considered it would be misleading simply to say "he is lying" for this falsely suggests that the beggar has children all or most of whom are well. Consequently an explanatory emendation, such as "He has no children at all," is necessary to avoid misunderstanding.

Although Strawson's emended account avoids the difficulties I have suggested with his earlier account, nevertheless, I have been inclined to treat the beggar's lie differently. Rather than say he lied (and told a lie) and thereby that what he said was false, thereby using the terms "lie" and "false" in a secondary or atypical way, I want to say that the beggar did *say* something which he correctly believed to be false, namely that he has children who need medical attention. To say that there is a presupposition unfulfilled seems unduly obscure, for that suggests, I think misleadingly, that we or he presuppose that he has children. Rather, in the context in which he utters those words, we can describe him as *saying* that he has children who are in need of medical attention. And under this description what he said was false, and not in some secondary sense of "false."⁹

6. *He must communicate.* Suppose *A* says to *B* what he believes or knows to be false in an attempt to deceive *B*. But suppose that *B* does not hear what *A* said. Has *A* lied to *B*? Has *A* told *B* a lie? It seems that we could say either, but to say that *A* has lied to *B* when *B* did not hear what *A* said, sounds a bit deviant. Certainly such a locution would be understood but so would we understand a foreigner who, in offering an apology said, "I am apologetic to you for what I did." It seems closer to the mark to say that *A* told a lie but *B* did not hear it; or that *A* was telling a lie to *B* but *B* did not hear it; or, perhaps that what *A* said to *B* was a lie but *B* did not hear it. Or should

we say, "*A* was trying or attempting to lie to *B* but *B* did not hear it?"

If *B* were deaf or dead, and *A* did not know this, the point would be slightly different. For in that case it seems quite queer to say "*A* lied to *B* but *B* could not hear because he was deaf," or "*A* was lying to *B* but *B* was dead." It is odd here even to say "*A* told *B* a lie but *B* was deaf (dead)." But we might say that what *A* said was a lie, but *B*, to whom the lie was addressed, was deaf (dead).¹⁰

Is telling a lie similar to issuing a warning? We could say that *A* issued a warning, but *B*, to whom the warning was addressed, was deaf or dead. Here it would be a bit odd to say that *A* issued a spoken warning to *B* but *B* was deaf or dead. But clearly it would be wrong to say that *A* warned *B* but *B* was deaf or dead. The point would be less clear, but nevertheless it might hold, when *B* simply did not hear. Even here it seems better to say that *A* tried to warn *B* but *B* wasn't listening.

The logical principle or tendency here seems to be the following: In order to complete the act of issuing a warning it is not necessary that the person to whom it was directed hear and understand it. But, to complete the act of warning it is necessary that the person to whom it was directed hear and understand it.

Should we say that to complete the act of uttering and possibly telling a lie the person to whom the lie was directed need not hear it; and to complete the act of lying the person to whom the lie is directed must hear and understand it? If this is at all plausible then the act of lying to *B* requires *B*'s hearing it, and the act of telling *B* a lie is a border line case and perhaps requires that *B* hear it. And, if this is so, then clearly it would be so when *B* is deaf or dead. Consequently, the clearest point would be that *A* can utter a lie, and what *A* said can be a lie, when *B* is deaf or dead; but *A* cannot lie to *B* when *B* is deaf or dead.

But then, how close is the analogy between "lie" and "warn"? "Warn" takes a direct object, and the verb phrase "issued a warning" does not take a direct object, but an indirect object. In this sense the verb "warn" is the name of an action which requires an object for making sense of the action. This suggests that for the act of warning

⁸ "A Reply to Mr. Sellars," *The Philosophical Review*, vol. 63 (1954), pp. 216-231.

⁹ Compare: "In the case of deliberate deception . . . the word 'false' may acquire a *secondary* use, which collides with the *primary* one." *Ibid.*, p. 230.

¹⁰ It must, of course, be assumed that *B* cannot read *A*'s lips, or the like.

to be completed there must be a person who is warned. Again, in issuing a warning, one must have someone in mind, for it makes no sense to say "I issued a warning, but I had no idea or intention of issuing it to anybody at all." And although a warning may not reach its intended or any unintended destination it was, nevertheless, issued. Now "lie" does not take a direct object; both "lie" and "told a lie" take indirect objects. By parity of reasoning neither should require that the lie come off as intended or be passed on to the intended or to any party. And if that is so, then the analogy with "warn" will not help to support the distinction between lying and telling a lie.

On the other hand, consider the verb "promise." "Promise" takes a direct object in "I promised three beers to the winner" or "I promised my daughter to him," and an indirect object in "I promised him that I would come." In no case does "promise" take the person to whom the promise was made as a direct object. Nevertheless, it does not seem quite right to say that *A* promised *B* that he would come, but *B* didn't (couldn't) hear what *A* said. For it may be that here, as well as with the verb "apologize" the act is not completed without communication, and thereby the description of the act could not be "*A* promised" or "*A* apologized." If this is clearer in the case of "promise" than in the case of "apologize,"¹¹ it might be mirrored in the fact that there is not a need for a preposition in "*A* promised *B*," but there must be a preposition in "*A* apologized to *B*." Consequently, the presence or absence of a direct object, whether the verb is called "transitive" or not, is not conclusive; but it may give a clue as to whether the action specified by the verb is one which (as grammarians say) "passes over to an object" and whether the action must pass over into an object. Again, in some cases it seems that the actions specified by some verbs which do not take direct objects cannot be completed if there is no communication to the person specified by the indirect object. On the other hand, "invite" takes a direct object (and, of course, does not require a preposition), but that

the act of inviting is incomplete without communication is not at all more obvious than that the act of promising or apologizing is incomplete without communication.

Certain general conclusions arise from all of this. I wondered which of the six features were necessary conditions for the application of which of ten locutions.

Saying something or doing something which could be described as saying something (Sect. 1), seemed to be a necessary condition for the application of any of the locutions. And knowing that what he says is false (Sect. 4), seemed necessary for none. Intending to deceive (Sect. 2), and believing that what he said is false (Sect. 5), both seemed necessary for "*A* was lying (to *B*)," (1, 2); and "*A* lied (to *B*)," (5, 6), but no others. Saying what is false (Sect. 3), seemed necessary for only "*A* was telling a lie (to *B*)," (3, 4); and "*A* told a lie (to *B*)," (7, 8); and "What *A* told (to *B*) was a lie" (9, 10). Communication (Sect. 6), with some reservations, seemed to divide the locutions differently. It seemed necessary for 2, 4, 6, 8, 10, but not for 1, 3, 5, 7, 9.

There seems to be call for another division, namely, between 1, 2, 3, 4, and 5, 6, 7, 8, 9, 10 (or between the progressive and simple verb forms).¹² Suppose *A* faints, is shot, or is similarly interrupted in the midst of speaking to *B*. The failure to complete what he was saying could make it wrong to say that he had lied (to *B*) or told a lie (to *B*), but granted the presence of other features, we could say that he was lying (to *B*) or telling a lie (to *B*) when he was interrupted. The interruption could come from natural forces or from human verbal or non-verbal intervention. On the other hand, *A* might have second thoughts in the course of his lying or telling a lie, in which case, if he stops soon enough (which point depends on particulars), it would be false to say that he lied or told a lie, but rather, given other features, that he was telling a lie or lying and changed his mind. This distinction, then, is between completing the act of lying or telling a lie and being in the course of performing the act.

University of Chicago

¹¹ "*A* apologized" is more likely than "*A* promised" when: (1) *B* wasn't listening, or (2) *B* didn't understand him because (i) it was in Latin, (ii) the noise was too great.

¹² The form "What he was telling (to *B*) was a lie" would go with 1, 2, 3, 4.

IV. DUTIES, RIGHTS, AND CLAIMS

JOEL FEINBERG*

AMONG the questions that still divide philosophers who are concerned with problems about rights are (1) whether, or to what extent, rights and duties are logically correlative, and (2) whether it is theoretically illuminating generally, and in particular, whether in considering question (1) it is strategically useful to treat rights as *claims*. Although question (1) is in a familiar sense a logical question (Do statements of duties entail statements of other people's rights, and do statements of rights entail statements of other people's duties?), this paper is more a descriptive or impressionistic study than a formalistic one. Part I consists of an examination of the many kinds of normative relations called "duty" with the aim of distinguishing those that are clearly correlated with other people's rights from those that apparently are not. The second part of the paper shifts the focus to rights and argues that there is at least one kind of talk about rights-as-claims that is neither reducible to, nor in any clear logical relation with, talk about duties. The word "claim" of course is ambiguous. Claims *to* (I shall argue) are not always expressible as claims *against*, and "having a claim to . . ." "making claim to . . ." are different sorts of things from "claiming that. . ." The paper concludes, however, that each of the ideas capable of being expressed by the word "claim" is essential either to the understanding or the just appreciation of rights.

I. DUTIES AND RIGHTS

Which of the various kinds of duty are necessarily correlated with the rights of other people? Consider first the relation between a debtor and his creditor. Indebtedness is the clearest example of one person *owing* something to another; and owing, in turn, is a perspicuous model for the interpretation of that treacherous little preposition "to" as it

occurs in the phrase "obligation *to* someone." Now it is unquestionably true that when one party *owes* something to another, the latter has a *right* to what he is owed. The debtor's obligation is his creditor's right seen from a different vantage point. A *duty of indebtedness*, moreover, entails a right of a very specific kind, called, in the jargon of jurisprudence, a positive *in personam* right, that is, a right against one specific person requiring him to perform a "positive act," not a mere omission.

A second class of duties, being based on promises, is also more properly called "obligations,"¹ but we can call them (other) *duties of commitment*. In discussing these, we must not be misled by the preposition "to." When a debtor owes money to a creditor, he can be said to have an obligation *to* the creditor; but the preposition here is ambiguous and obscures the distinction between two different offices the creditor occupies. On the one hand he is the one to whom the obligation is *owed*, and the one therefore who can claim it as his due. On the other hand, he is the intended beneficiary of his debtor's promised act. This dual role is also played sometimes by persons owed other kinds of duties of commitment. If Abel promises Baker to meet him at a certain time, or to shine his shoes, or favorably review his book, then Baker is both claimant and intended beneficiary of Abel's duty. There may, of course, be others who stand to gain, if only indirectly, from Abel's discharge of his obligation, but in most cases these so-called "third party beneficiaries" will profit in merely picayune and remote ways.

Sometimes, however, there is a separation of offices, and the intended direct beneficiary is not the promisee, but instead some third party designated by the promisee. This class of transactions can be further subdivided: In some cases, only the promisee is the claimant, or right-holder, while in other cases, both the promisee and the

* This paper is a slightly expanded version of my contribution to a Symposium on Human Rights at the meeting of the Eastern Division of the American Philosophical Association, December 27, 1964.

¹ For subtle discussions of the distinction between obligations and duties, see E. J. Lemmon, "Moral Dilemmas," *The Philosophical Review*, vol. 71 (1962), pp. 139-158; and Richard B. Brandt, "The Concepts of Obligation and Duty," *Mind*, vol. 73 (1964), pp. 374-393.

third-party beneficiary have rights to the promisor's performance. If Abel promises Baker to look after Baker's dog Fido, then Fido is the direct beneficiary of the promised services, while Baker himself, and probably only Baker, is the claimant. On the other hand, if Baker designates his wife or mother (or dog?) as beneficiary on his life insurance policy, then *both* Baker as promisee and the designated beneficiary can be said to have a right that the benefit-payment go to the beneficiary. (Even after Baker is dead, it can be said that the insurance company *owes it to him* to pay the beneficiary.)

Philosophers² have sometimes found it useful to direct their attention primarily to those cases where third-party beneficiaries do *not* derive rights from promises, for such cases illustrate most clearly that promisee and intended beneficiary *are* distinct moral offices. Furthermore, in cases of that kind, the distinction between the two offices is not totally obscured by ordinary language which distinguishes (sometimes) between duties *to* claimants and duties *toward*, or in respect to, beneficiaries. But total preoccupation with this kind of case is as dangerous as it is unnecessary. The danger is that it might blind us to the large class of cases where third-party beneficiaries, both in law³ and morals, do have a claim-right against the promisor; and in any case, the distinction between the offices of promisee and beneficiary can be equally well made out in the case where the third-party beneficiary and the promisee are both claimants. For in these cases, it does not *follow necessarily* from the fact that a person is an intended beneficiary of a promised service that he has a right to it, whereas it always follows necessarily from the fact that a person is a promisee that he has a right to what is promised. This difference, of course, would not be possible if beneficiary and promisee were not distinct offices. Another way of putting the point is to say that the rights of third-party beneficiaries, unlike those of promisees, are not logically correlated with the obligations of the promisor, but correlated only in virtue of moral or judicial policies and rules. In those cases where there is

some temptation to say that the right of a third-party beneficiary is logically correlated with a promisor's duty, the temptation will be at least as great to say that the promisor made a *tacit promise* to the beneficiary in addition to the express promise to his promisee. I have in mind those cases where the parties to the promise allow the promise to be known to the third-party beneficiary, and the latter acts in reliance on its performance.

In all of these cases, the important relation for our present purposes is that between promisor and *claimant*, whether the latter be promisee, or beneficiary, or both. This relation is another proper and familiar case of owing, although it is already one step removed from what we might call the "paradigm case" of indebtedness. Duties of commitment, like the standard cases of owing, are obviously correlated with other people's *in personam* rights. The claimant has a right *in personam* against the promisor to either a positive performance, as in the case of feeding Fido, or else a forbearance, as when I promise to waive my right to keep you off my land, giving you thereby a claim to my non-interference, that is, a negative *in personam* right.

Similar remarks can be made about a third class of duties, the *duties of reparation*. When your loss is "my fault," that is, when it was caused by my negligence, recklessness, impulsiveness, carelessness, dishonesty, malevolence, or the like, then I have a duty to you to repair the harm or otherwise make good the loss. I "owe" reparation to you in much the same manner as I would owe you the return of something I borrowed or took from you without your permission. My duty, in these examples, is to return to you what is really your own, or where this is impossible, something of equivalent value; and your correlative right is a claim *in personam* to my positive services.

A fourth class of duties also permits talk of "owing" something to someone, although we are now at least two steps away from the example of indebtedness. "Mr. Churchill feels that he owes this legacy to the world," said a 1964 advertisement for a set of recordings by Winston Churchill of public and private speeches, letters, and remini-

² Most notably, H. L. A. Hart, "Are There Any Natural Rights?" *The Philosophical Review*, vol. 64 (1955), pp. 179-182.

³ For a subtle and detailed discussion of this complex topic see *Corbin on Contracts* (St. Paul, Minn., West, 1952), pp. 723-789. Corbin writes on p. 733: "The following is an attempt at a consistent statement of the generally prevailing law: A third party who is not a promisee and who gave no consideration has an enforceable right by reason of a contract made by two others (1) if he is a creditor of the promisee or of some other person and the contract calls for a performance by the promisor in satisfaction of the obligation; or (2) if the promised performance will be of pecuniary benefit to him and the contract is so expressed as to give the promisor reason to know that such benefit is contemplated by the promisee as one of the motivating causes of his making the contract. A third party may be included within both of these provisions at once, but need not be. One who is included within neither of them has no right, even though performance will incidentally benefit him."

scences. Presumably, Sir Winston did not feel that he must simply return what he had borrowed or keep some sort of promise, express or implied. I suspect rather that he felt a duty to give to the world something that it needs, but which he, at age 90, no longer had reason to keep his exclusive possession. I propose to call this, and other more worldly examples of the duty abundance owes to need, *duties of need-fulfillment*. Such duties clearly give rise to positive *in personam* rights, often in many claimants.

A fifth class of duties is related to gratitude, but had better be called *duties of reciprocation*, since gratitude, a feeling, is a less appropriate subject for duty than reciprocation, which is, after all, action. There are, moreover, other confusions commonly infecting the idea of a "duty of gratitude." Many writers speak of duties of gratitude as if they were special instances, or perhaps informal analogues, of duties of indebtedness. But gratitude, I submit, feels nothing at all like indebtedness. When a person under no duty to me does me a service or helps me out of a jam, from what I imagine to be benevolent motives, my feelings of gratitude toward him bears no important resemblance to the feeling I have toward a merchant who ships me ordered goods before I pay for them. The cause of the widespread confusion of gratitude with moral indebtedness, I suspect, is a disposition, allegedly characteristic of but certainly not peculiar to the Japanese, to feel some loss of status when helped by others, and some consequent resentment of the benefactor under the respectable mask of "gratitude." We feel impelled to pay back a benefactor sometimes because we feel that his benefaction has made him "one up" and we want to get even.

The expression "duty of reciprocation" is better used for a different kind of case: My benefactor once freely offered me his services when I needed them. There was, on that occasion, nothing for me to do in return but express my deepest gratitude to him. (How alien to gratitude any sort of *payment* would have been!) But now circumstances have arisen in which he needs help, and I am in a position to help him. Surely, I *owe* him my services now, and he would be entitled to resent my failure to come through. In short, he has a right to my help now, and I have a correlative duty to proffer it to him. Like the other examples, the right in this case is *in personam* and typically positive.

I think I have now enumerated the main classes of duties that permit talk of one person owing something to another. Of course, there may be a

very wide sense of "owe" in which it goes with all talk of duty, perhaps as a kind of synonym for the feeling of requirement or "must do" that goes with all duty. But still in respect to the remaining classes of duties, while one must do something, this is not because he *owes* it to someone to do it.

The sixth kind of duty is typified by the duty we all have to stay off a landowner's property. I don't think we would naturally speak of this duty as something "owed" to the landowner, although I admit the law doesn't hesitate to speak that way. In acknowledging a duty not to interfere with another's property, we show our respect for his interest in the exclusive possession and control of it. Such duties of noninterference with the person or (prototypically) the property of another, I propose to call *duties of respect*. This use of the word "respect" is not the only one, but it is, I think, a familiar one. Webster's dictionary puts it thus, "... to esteem; value; hence to refrain from obtruding upon or interfering with; as to respect a person's privacy."

The rights correlative with duties of respect are typically negative, that is, rights to other people's abstentions, forbearances, or noninterference, and unlike the rights discussed in our earlier examples, they are what lawyers call *in rem* rather than *in personam* rights. An *in rem* right holds, not against some specific namable person or persons but rather, in the legal phrase, against the world at large. In saying that "the whole world" has a duty to stay off my land, all I can mean, of course, is that any person in a position to enter my property has a duty to stay out. That implies that even General De Gaulle, if I wished to keep him out, would have to stop at my gate. My right *in rem*, in imposing on others a duty of respect, is itself no respecter of persons.

Are all *in rem* rights negative? There is no denying that negative *in rem* rights, modeled after the proprietary right and then extended to cover personal interests as well, have had an enormous influence on political thought, especially in America. They dominated lists of "natural rights," for example, in various eighteenth-century manifestos. Still there are positive *in rem* rights too, whose importance has come to be appreciated anew only in recent decades. Consider, for example, the duty of care that every citizen is said to owe to any and every person in a position to be injured by his negligence. I have this duty to some degree even to the uninvited trespasser on my land. Or consider the duty (not equally recognized in our law) that every citizen has to come to the aid of

accident victims. These unfortunates have a right to be assisted that holds against every or any person in a position to help. I propose to call such positive *in rem* rights, *rights of community membership*, because it is their recognition, more than anything else, that molds a society into a cohesive community.

An eighth class of duties, which I shall call *duties of status*, is perhaps the original from which many of the others derive. In the Middle Ages, a "duty" was something *due* a feudal lord, in virtue of his role and its status in the social system, from one of his inferiors, a vassal or a serf, in virtue of *his* status. A person, in being born into his relatively fixed position in the social order, was at the same time born into the duties that went with, and indeed defined, that position. One's duty was conceived as a kind of payment of one's proper share to the general economy of interests, and of course there were different shares to be exacted from different ranks and stations. Doing one's duty might be paying in crops or livestock, or performing assigned tasks at periodic intervals, or for the higher ranks contributing troops, horses, and weapons. Very likely these payments were made in a spirit similar to that in which club members pay their dues, especially in a club whose rules prescribe different types of payment for different types of members.

It was not difficult, in a rigidly hierarchical society, to know *to whom* one's duty was owed, for payments were generally from lower rank to higher, with the occupant of the higher rank always capable of exacting payment, if necessary, by force. With the decline of feudalism, however, it became increasingly difficult to find a specific claimant for every status duty. Offices and roles, of course, still survived, and carried with them their attached duties, but there was no longer a single clear line of direction in which they were owed, or single source of sanctions for their enforcement. To be sure, later when contract came to supplant status as a primary principle of social organization, one could in theory come to think of the duties of his job as derived from the "employment contract" and therefore *owed as obligations* to the boss, as promisee. This was, however, seldom a convincing myth. Employment contracts were often unfairly bargained, and by the time conditions improved

in that respect, the employer was so vast and impersonal he could hardly be conceived as the claimant of a personal obligation. Hence, duties of status have come less and less to be thought of as *owed* to anyone.

The concept of a duty, however, has by no means completely forgotten its past. It still preserves its ancestral connection with offices, stations, and jobs;⁴ it is still bound up, however remotely, with the idea of coercion, and it still commonly suggests the idea of a fair share of burdens, imposed on one as a levy, for the promotion of socially shared interests. In group undertakings, it is often said that "if only everybody pitches in and does his *share*, the job will be done." The share we are thereby exhorted to contribute is, of course, the very same as our duty, and it will be greater for the rich than the poor and lesser for the weak than the strong.

Does it still make sense to ask *to whom* one's duty of status is owed? Perhaps, but we can no longer always expect a simple answer mentioning some specific person, such as "one's feudal lord," or "one's employer." To whom does the left tackle on a football team owe his assigned duty to block the player opposing him? In a case like this it is odd to say that the duty is owed to anyone but "the team." And similarly we often hear of status-duties owed "to the company," or "to the university," or "to one's country." And in still other cases, for example the duty of a janitor to sweep the corridors, it might plausibly be urged that the duty is owed to *no one at all*, although it is no less a duty for that.

Perhaps the most important feature of our talk about duties I have only mentioned up to now, and that is the alliance of the idea of duty with the idea of coercion. A duty, whatever else it be, is something *required* of one. That is to say first of all that a duty, like an obligation, is something that *obliges*. It is something we conceive of as *imposed* upon our inclinations, something we must do *whether we want to or not*. Second, a requirement is, in a perfectly good sense, a *liability*, something *we must do or else* "face the consequences" (punishment, firing, guilt feelings). When the coercive element common to duties and obligations is in clear focus, it is likely to seem so centrally impor-

⁴ If the phrase "moral duty," unlike "moral obligation," sounds odd in our ears, it is, I submit, because it suggests that there is an office or job of "man as such"—a most dubious metaphysical idea (*pace* Plato and Aristotle). Generally speaking, it is difficult to find plausible analogies between our moral problems as men and our moral problems as office holders. There are no analogies of this kind to compare with the close analogy between our obligations as contract signers and our moral commitments as promisors. Hence, the appropriateness of the phrase "moral obligation" and the oddness of "moral duty."

tant as to dim the various differences between the two conceptions, as when the lawbooks, for example, speak interchangeably of imposing duties and obligations. Moreover, both terms, "duty" and "obligation," have developed extended senses in which *only* the coercive feature is essential, as when we speak, for example, of an action, fitting perhaps as a "gesture," or symbolic expression of feeling, as a duty when it seems to have a "compelling" appropriateness. "Duty" and "obligation" both tend now to be used for any action we feel we must (for whatever reason) do.

Duties of compelling appropriateness are perhaps only duties in an extended sense, but still there is no harm in labeling them and including them in our catalogue.⁵ The class probably includes such philosophically puzzling specimens as "duties of perfection," "duties of self-sacrifice," "duties of love," "duties of vicarious gratitude," and so on. It is clear, I think, that people who feel that they have duties of this kind do not feel them as owed to anyone.⁶

In speaking of a duty as a liability, we should take care to distinguish it from another kind of liability also imposed by roles and jobs, namely those that have come to be called *responsibilities*. A responsibility, like a duty, is both a burden and a liability; but unlike a duty it carries considerable discretion (sometimes called "authority") along with it. A goal is assigned and the means of achieving it are left to the independent judgment of the responsible party. Moreover, the liability to unwanted consequences in the case of a responsibility tends to be "stricter" than in the case of a mere duty. That a man tried his best is more likely to be accepted as an excuse for failure to perform one's duty than for failure to fulfill one's responsibility. Indeed, the more discretion allowed in the

responsibility assignment, the stricter the liability for failure is likely to be. In general, the closer the resemblance of a task assignment to the purely non-discretionary cases, where for example, the officer's command "Fire!" imposes the duty to pull the trigger, or where the annual dues notice imposes the duty to pay, the more likely we are to characterize it as a duty, and the less likely to call it a responsibility. A "duty to obey" makes sense; but there could be no such thing as a "responsibility to obey."⁷

This leads us to our final class of duties, the *duties of obedience*. The medieval lord was, in relation to his serf's duty, both beneficiary, claimant, and enforcing authority. In the complication of social roles that followed the collapse of feudalism, the separation of these three offices became common. In particular, the man who can, in some institutions, command the performance of duty from us, and back up his command with sanctions, is not always the same as the person, if there is one, to whom that duty is *owed*. It appears then to be a quite different sense of the preposition "to" in which we have duties *to* a commanding authority. And yet we commonly enough hear talk of "owing obedience" to parents, police officers, and bosses, and these authorities speak readily enough of having a claim to our obedience. Does an authority then have a *right* to be obeyed by his inferiors?

A traffic cop blows his whistle, points, and shouts "Stop!" This, of course, imposes a (legal) duty on a motorist to stop. Still it is not true that the policeman can claim the motorist's stopping as *his* due or that the motorist owes it to *him* to stop. Perhaps the policeman has an *official* right, derived from his status *qua* policeman, rather than a *personal* right, that the motorist stop. I suspect, however, that this is simply a roundabout way of

⁵ It would be going too far, however, to include "duties of compelling attractiveness," which could be duties in no proper sense, paradigmatic or extended.

⁶ H. B. Acton gives several examples of persons who act "on a conception of duty that requires [them] to give benefits to others much in excess of what is [believed to be] their right." "Thus Celia in Eliot's *The Cocktail Party* must have considered she had duties to savages who certainly had no right to services from her. The man who, in Malraux's novel, gave all his supply of poison to his fellow-prisoners to enable them by suicide to escape the burning alive which was to be their fate and his, probably did not think that they had more right to the poison than he had, though he thought it his duty to give it to them. Some of these supererogatory acts may be a form of disguised egoism, the agent regarding himself as worthy of a much stricter code of behaviour than the majority of people. Others are the result of compassion or benevolence, and in that way, perhaps outside the sphere of rights and duties. But some of the more impressive acts of moral heroism appear to be performed at the behest of an exacting sense of duty without their being [I should add 'any sense of'] corresponding rights on the part of the beneficiaries." H. B. Acton, "Symposium on 'Rights,'" *Proceedings of the Aristotelian Society*, Supplementary Volume 24 (1950), pp. 107-108.

⁷ The title of a current paperback book found in most drugstores is *The Sexual Responsibility of Woman*. The book spends several hundred pages describing the many kinds of situations in married life that call for the exercise of intelligence, judgment, and adaptability on the part of the wife. On the other hand, there could be no doubt what might be meant by a book, published say in Victorian England, with the title *The Sexual Duty of Woman*, for the sexual duty of a wife, if there could be such a thing, could only be to submit. There could be no such thing as a "responsibility to submit."

saying that the policeman's office confers on him the authority to command motorists to stop, which of course is beyond question, yet does nothing to settle the further question whether authorities can be said to have a right that persons do as they command. In any case, many duties of obedience are "owed" to impersonal authority like "the law," or a painted stop sign. Here it is especially difficult to find an assignable person who can claim another's stopping as his due. Some duties of obedience, then, seem to entail no correlative rights; and if my suspicion is correct, none of them do. For if the preposition "to" in the phrase "duty (of obedience) to one's superior" means the same as it does in the expression "answerable to so-and-so for his failure," and I suspect that this is so, then the authority to whom one "owes" obedience is not a *claimant* in the manner of (say) a creditor, but rather simply the one who may properly command performance of duty and apply sanctions in case of failure. The little preposition "to" then is triply ambiguous when used with "duty." One can have a duty *to* his claimant, *to* (or toward) a mere beneficiary, and be liable for failure *to* an authority; but it is only the claimant who can properly be said to have a right to one's performance.

In summary, duties of indebtedness, commitment, reparation, need-fulfillment, and reciprocity are necessarily correlated with other people's *in personam* rights. Duties of respect and community membership are necessarily correlated with other people's *in rem* rights, negative in the case of duties of respect, positive in the case of duties of community membership. Finally, duties of status, duties of obedience, and duties of compelling appropriateness are not necessarily correlated with other people's rights.

II. RIGHTS AS CLAIMS TO . . .

Having described the various kinds of duties that *are* correlated with rights, have we thereby done

all that is necessary to elucidate the concept of a right? Many writers seem to think so.⁸ I am inclined, however, to agree with Richard Wasserstrom⁹ that we have not until we have said something further about rights as *claims*. It will not help to attempt a formal definition of rights in terms of claims, for the idea of a right is already included in that of a claim, and we would fall into a circle. Nevertheless, certain facts about rights, more easily, if not solely, expressible in the language of claims and claiming, are necessary to a full understanding of what rights are and why they are so vitally important.

There may at first sight be grounds for holding that claims are always *against* someone, and therefore necessarily correlated with the duties of those against whom they hold; but there is a sense of "claim," closely related to "need," in which this is not always so. Imagine a hungry, sickly, fatherless infant, one of a dozen children of a desperately impoverished and illiterate mother in a squalid Mexican slum. Doesn't this child have a *claim* to be fed, to be given medical care, to be taught to read? Can't we know this before we have any idea where correlative duties lie? Won't we still believe it even if we despair of finding anyone whose duty it is to provide these things? Indeed, if we do finally *assign* the duty to someone, I suspect we would do so because there *is* this prior claim, looking, so to speak, for a duty to go with it.

In our time it is commonplace to speak of *needs* as "constituting claims." William James thought that every interest is a kind of claim against the world and that the validity of an interest *qua* claim lies "in its mere existence as a matter of fact."¹⁰ This probably goes too far. We don't think of every desire or even every need as a claim, but important needs are another matter. They "cry out," we say, for satisfaction. (Note the etymological connection of "claim" with "clamor.") And when they cry no proper name but only their own need, we speak of their claims "against the world"; but this is but a rhetorical way of saying "claim against no one

⁸ Howard Warrender speaks of rights as "merely the shadows cast by [other people's] duties" (*The Political Philosophy of Hobbes* [Oxford, Clarendon Press, 1957], p. 19); S. I. Benn and Richard Peters write that "Right and duty are different names for the same normative relation, according to the point of view from which it is regarded" (*Social Principles and the Democratic State* [London, Allen & Unwin, 1959], p. 89); and Richard Brandt writes that "a society with a language that had no term corresponding to 'a right' might still be said to have the *concept* of a right, if it were recognized that people have the *obligations* toward others which are the ones that correspond with rights" (*Ethical Theory* [Englewood Cliffs, N.J., Prentice-Hall, 1959], p. 441). In his sensitive discussion, Professor Brandt does allow, however, that there are important differences in emphasis and "overtone" between talk of rights and equivalent talk of duties.

⁹ "Rights, Human Rights, and Racial Discrimination," *The Journal of Philosophy*, vol. 61 (1964), pp. 628-641.

¹⁰ William James, "The Moral Philosopher and the Moral Life," *International Journal of Ethics* (1891), reprinted in *Essays in Pragmatism*, ed. A. Castell (New York, Hafner, 1948), p. 73.

at all." (Or perhaps a "claim against the world" is like an explosion in the desert—there is no one to hear it, but were anyone to get close to it what a commotion he would hear, and what an impact he would feel! So it is perhaps with my little Mexican urchin. Perhaps her claim is like a "permanent possibility of sensation," real enough, though no one comes within its range. Still note what one does hear, if he is not morally deaf, when he comes close enough: He hears a *crying need*, a claim to . . . that is so strong it may be felt as a claim *against*. . .)

The right to education, like other positive *in rem* rights peculiar to twentieth-century manifestos, has caused much confusion and dissension, partly because theorists, in their eagerness to provide schematic translations for all rights in terms of other people's duties, have simply overlooked the sense of "right" uppermost in the minds of manifesto writers. Professor Brandt, for instance, says of "my having a right to an education" that it "implies roughly that each individual in my community has an obligation to do what he can [in another formulation Brandt says "to cooperate substantially"] in view of his opportunities and capacities and other obligations, to secure and maintain a system in which I and persons in my position are provided with an opportunity for education."¹¹ But surely there is a familiar sense of "right" that requires more than that others try, or "do what they can" (considering of course how *busy* they all are) or "cooperate substantially." My right in this sense is *to* the education (there is that preposition again in still a fourth sense) and not simply to other people's dutiful efforts. More likely my right in this case (if I have one) entails not simply a duty to try but a responsibility to succeed; but even this doesn't do the whole job of

translation, for there is a **MUST HAVE** here not wholly translatable into any number of **MUST DO's**.¹²

III. CLAIMING THAT ONE HAS A RIGHT

I wish finally to emphasize the importance of the verb "to claim," not to the analysis of a right, but to an understanding of why rights are, in Wasserstrom's phrase, such "valuable commodities."¹³ To claim that one has a right (or for that matter that one has any of the other things one might claim—knowledge, ability, whatever) is to *assert* in such a manner as to demand or insist that what is asserted be *recognized*. It is my contention that for every right there is a further right to claim, in appropriate circumstances,¹⁴ that one has that right. Why is the right to demand recognition of one's rights so important? The reason, I think, is that if one begged, pleaded, or prayed for recognition merely, at best one would receive a kind of beneficent treatment easily confused with the acknowledgment of rights, but in fact altogether foreign and deadly to it.

There are in general two quite distinct kinds of moral transaction. On the one hand there are gifts and services and favors motivated by love or pity or mercy and for which gratitude is the sole fitting response. On the other hand there are dutiful actions and omissions called for by the rights of other people. These can be demanded, claimed, insisted upon, without embarrassment or shame. When not forthcoming, the appropriate reaction is indignation; and when duly done there is no place for gratitude, an expression of which would suggest that it is not simply one's own or one's due

¹¹ Brandt, *Ethical Theory*, *op. cit.*, p. 437.

¹² No doubt this is an extended sense of "right." I insist only that it is a proper and important one. Note that there are parallel extended senses of "claim" and "demand" both in quite general circulation. Webster's gives as its fourth sense of "demand," for example, "to call for or require as necessary or useful; to be in urgent need of, as in the phrase 'the case demands care'." It is in this sense, *at the very least*, that children require education, sickness calls for medicine, and hunger demands food.

¹³ Wasserstrom, *op. cit.*, p. 629.

¹⁴ G. J. Warnock in a useful article ("Claims to Knowledge," *Proceedings of the Aristotelian Society*, Supplementary Volume 36 [1962], p. 21) says that when I claim to others that I *know* something, I am not merely asserting it, but rather I am "obtruding my putative knowledge upon their attention, demanding that it be recognized, that appropriate notice be taken of it by those concerned. . . ." This sounds like the behavior of a perfect boor! I don't wish to contend that all rights confer the additional right to be boorish, but only that one may insist, in the appropriate circumstances and with only an appropriate degree of vehemence, that one's right be recognized.

A list of "appropriate circumstances" would include occasions when one is challenged, when his rights are explicitly denied, when he must make application for them, where his possession seems insufficiently acknowledged or appreciated, etc. There may even be appropriate circumstances for one's demanding recognition of his second-level right to claim ground-level rights; but circumstances would rarely be appropriate for claiming third-level rights, and probably never for levels higher than that. So the contention in the text that for every right there is a right to claim *in appropriate circumstances* that one has that right leads to no kind of vicious regress.

that one was given.¹⁵ Both kinds of transaction are important, and any world with one but not the other would—in Wasserstrom's phrase—be “morally impoverished.” A world without loving favors would be cold and dangerous; a world full of kindness, but without universal rights, would be one in which self-respect would be rare and difficult. Too much gratitude is a very bad thing, leading donors to be complacent and hypocritical,

and doing worse harm still to the recipients. If the rugged individualist who boasts in his blindness that he owes nothing to any man is no moral paragon, neither is he who feels gratitude for everything, for that is a kind of self-abasement; and from men who respect not their own interests nor feel even their most basic needs as claims, little good, and probably considerable mischief, can be anticipated.

Princeton University

¹⁵ The obverse of this point is worth noting too. Gratitude often is the appropriate response to a person's deliberate *failure* to press for his rights. I quote from a perceptive editorial in *The New Yorker* (June 3, 1961, p. 23): “... Conceivably it is in the national interest to persuade Negro leaders to set a slower pace, but the argument is one that does not permit a high moral tone. One can hardly, with justice, inform a Negro that he has a duty as a citizen to refrain from sharing in the rights of citizenship. We can imagine asking it, under special circumstances, but only as the immense favor that it would be, rather than as the obligation it certainly is not.”

V. McTAGGART'S ANALYSIS OF TIME

RICHARD M. GALE

McTAGGART'S famed argument for the unreality of time, first presented by him in 1908 in *Mind*, comprises both a positive and a negative thesis. The positive thesis, which is presented in the first part of the argument, contains an analysis of the concept of time, which McTaggart claims to be the only correct one. The negative thesis which is presented in the second part of the argument attempts to show that this analysis of the concept of time entails a contradiction. The assumption here is that any concept which is contradictory cannot be true of reality, and therefore time is unreal. It is vital to distinguish between these two theses, for, as we shall see, one group of analytic philosophers who answered his argument agreed with his positive thesis concerning the correct analysis of the concept of time, but they disagreed with his negative thesis that this analysis entails a contradiction. On the other hand, another group of analytic philosophers who refuted his argument disagreed with his positive thesis and claimed that time could be real even if his negative thesis is correct. Thus, while these two groups of philosophers are in agreement as to the unsoundness of McTaggart's argument, they adopt competing analyses of the concept of time in refuting this argument. The purpose of this paper will be to examine critically McTaggart's positive thesis and thereby bring into sharper focus the exact nature of the dispute between these two groups of analytic philosophers. But before doing this, a rough sketch will be given of McTaggart's argument as a whole.

In the first part of his argument McTaggart analyzes the concept of time in terms of two different types of temporal facts. First, there are facts about temporal relations of precedence and subsequence between events, and, second, there are facts about the pastness, presentness, and futurity of these same events. Corresponding to the first type of temporal facts is a series of events, called the "*B-Series*," which runs from earlier to later, its generating relation being *earlier (later) than*; while corresponding to the second type of temporal facts is a series of events, called the "*A-Series*,"

which runs from the past through the present and through the present to the future. While both of these series are essential for the reality of time, the *A-Series* is the more fundamental of the two, since the *B-Series* can be derived from it alone. The arguments advanced by McTaggart for the necessity and fundamentality of the *A-Series* will be examined shortly.

The second part of McTaggart's argument attempts to show that the *A-Series*, which has been shown in the first part to be essential for the reality of time, entails a contradiction and that therefore time is unreal. His main argument is as follows. Every event in the *A-Series*, assuming that there is no first or last event, has the mutually incompatible properties of being past, present, and future, which is a contradiction. The obvious reply to this seeming contradiction is to say that no event has two or more of these incompatible properties *at the same time*, but rather has them successively *at different moments of time*. But this reply will not do, since it involves either a vicious circle or a vicious infinite regress. What we have done is to explain away the contradiction of an event in the first-order time-series being past, present, and future by claiming that it has these determinations successively *at moments of time* in a second-order time-series. But since this second-order time-series is a time-series, the moments of time which are its members must have the properties of being past, present, and future. Therefore, we have explained away the contradiction inherent in the first-order *A-Series*, brought about by the fact that every event in it has all three mutually incompatible determinations of past, present, and future, by introducing a second-order *A-Series*. And this is to reason in a vicious circle, since we must presuppose an *A-Series* to rid the first *A-Series* of contradiction.

If we should try to remove the contradiction inherent in this second-order *A-Series* because all the moments of time in it are past, present, and future, by saying that these moments are successively future, present, and past at moments of time in some third-order *A-Series*, we are merely transfer-

ring this contradiction to this third-order *A*-Series. We are launched on a vicious infinite regress because at any point in this regress at which we stop we are left with a contradictory *A*-Series. The curse of contradiction pursues us down this long infinite regress, being a sort of baton that each *A*-Series passes on to the *A*-Series which is one-order higher than itself.¹

There are two different type of answers given to McTaggart's argument by analysts: (1) the *B*-Series alone is sufficient to account for time—the *A*-Series being reducible to the *B*-Series—and therefore the *A*-Series is not essential for the reality of time; and (2) the *A*-Series alone is sufficient to account for time—the *B*-Series being reducible to the *A*-Series—but the concept of the *A*-Series does not contain a contradiction. Answer (1), to be called the "*B*-Theory Answer," attacks McTaggart's positive thesis by denying that the *A*-Series is essential for time. It supports this contention by arguing that, contrary to what McTaggart said, the *A*-Series is reducible to the *B*-Series.² Answer (2), to be called the "*A*-Theory Answer," basically agrees with McTaggart's positive thesis that the *A*-Series is both necessary and fundamental, but denies his negative thesis that the concept of the *A*-Series is contradictory.³ We shall now critically examine McTaggart's analysis of the *A*-Series to show that there is a crucial obscurity in the concept of the *A*-Series which has gone unnoticed by proponents of both the *A*- and *B*-Theory Answers. Once the concept of the *A*-Series is properly clarified, it will be seen that the *A*-Series, *properly understood*, is primitive, the *B*-Series being reducible to it. However, this thesis is not nearly as exciting as the one which the defenders of the *A*-Theory Answer thought they were espousing. While my clarification of the concept of the *A*-Series will tend to trivialize the *A*-Theory Answer's thesis of the primitiveness of the *A*-Series as opposed to the *B*-Series, what it will indicate about our concept of time is far from trivial, and, in fact, is of the highest importance.

Let us now scrutinize McTaggart's analysis of

time. His analysis is phenomenological, being based on the way in which temporal positions *appear* to us, but everything he says could be recast, and for purposes of clarity needs to be recast, in a linguistic idiom which describes the different ways in which we *talk* about temporal positions. He begins by saying:

Positions in time, as time appears to us *prima facie*, are distinguished in two ways. Each position is Earlier than some and Later than some of the other positions. . . . In the second place, each position is either Past, Present, or Future. The distinctions of the former class are permanent, while those of the latter are not.⁴

This can be rephrased linguistically as:

There are two fundamentally different ways in which we make temporal determinations. First, we can say that one event is⁵ earlier (later) than some other event; and, second, we can say that some event is now past (present, future). The sentences employed in making claims of the first sort make statements having the same truth-value every time they are uttered, while the sentences employed in the second sort of temporal determination may make statements having different truth-values if uttered at different times.

These two different ways of experiencing or talking about time are supposed to determine two different time-series.

For the sake of brevity I shall give the name of the *A*-Series to that series of positions which runs from the far past through the near past to the present, and then from the present through the near future to the far future, or conversely. The series of positions which runs from earlier to later, or conversely, I shall call the *B*-series.⁶

McTaggart's concept of the *B*-Series is clear because we can see the connection between it and our use of the expression "earlier (later) than." This connection consists in the fact that *earlier (later) than* is the generating relation of the *B*-Series; that is, given any two non-simultaneous events, *x* and *y*, either *x* is earlier (later) than *y* or *y* is

¹ It is not the purpose of this paper to answer this argument. For a complete bibliography and summary of the answers given to this argument see L. Mink, "Time, McTaggart and Pickwickian Language," *The Philosophical Quarterly*, vol. 10 (1960), pp. 254-263.

² The most illustrious defender of this thesis is Bertrand Russell, and among his followers are R. B. Braithwaite, A. J. Ayer, W. V. Quine, N. Goodman, D. Williams, and J. J. C. Smart.

³ C. D. Broad is the leading exponent of this answer, and among his followers are J. Wisdom, S. Stebbing, D. F. Pears, and W. Sellars.

⁴ J. M. E. McTaggart, *The Nature of Existence*, vol. II (Cambridge, 1927), pp. 9-10.

⁵ We will adopt the convention in this paper of italicizing tenseless verbs and copulae.

⁶ McTaggart, *op. cit.*, p. 10.

earlier (later) than x . The relation *earlier (later) than* is connected in the set of all non-simultaneous events.

Unfortunately, McTaggart's concept of the *A*-Series is not as clear, for while we can see the connection between the *B*-Series and talk about earlier or later than, we do not see the connection between the *A*-Series and our use of the predicates "is present," "is past" and "is future." Since the *A*-Series is a series, it must have a generating relation. But what could it be? We know only that the generating relation of the *A*-Series, whatever it might be, must contain a temporal index since there is a different *A*-Series at any two different moments of time. The *A*-Series which McTaggart describes cannot be generated by the unqualified temporal indices "past," "present," and "future" since the *A*-Series involves not only that the events comprising it are past, present, and future but also that they are past and future by varying degrees. The necessity for the latter is contained in McTaggart's description of the *A*-Series as one "which runs from the *far* past through the *near* past. . . ." The use of the unqualified predicate "is now past" cannot distinguish between events in the far and in the near past: it determines a "blob past." Past, present, and future, rather than being the generating relation of the *A*-Series, merely indicate the kind of elements in the *A*-Series, just as +, - and 0 do not generate the series of integers but only indicate the kind of elements that belong to the series. Before attempting to answer the question as to what could be the generating relation of the *A*-Series, we shall consider McTaggart's arguments for the necessity and fundamentality of the *A*-Series.

McTaggart claimed that there could not be time without both a *B*-Series and an *A*-Series. His arguments in support of this are directed only to showing the necessity for the *A*-Series. Probably he thought that the necessity for the *B*-Series is so obvious as to require no further comment; for to conceive of a "time" which admitted of no distinctions between earlier and later times would be a conceptual absurdity.⁷ While philosophers have rarely doubted the necessity for there to be a *B*-Series if there is to be time, several of them, in

particular those who defend the *B*-Theory Answer, have questioned the necessity of the *A*-Series. They have argued that although we always experience events as forming both a *B*- and an *A*-Series the determinations of events as past, present, and future are delusory or purely psychological, so that only the *B*-Series is objectively real. Against this line of reasoning McTaggart advances two arguments, the first showing that change requires an *A*-Series and the second, which is also an argument for the fundamentality of the *A*-Series over the *B*-Series, showing that "earlier (later) than" can be defined in terms of "past," "present," and "future."

McTaggart's first argument for the necessity of the *A*-Series has as its basic premiss that there cannot be time without change, which would be granted by proponents of both the *A*- and *B*-Theory Answers to his argument. The purpose of this argument is to show that there could not be change (and therefore time) without the *A*-Series. The argument proceeds by examining every possible candidate for the title of "change" other than changes in the pastness, presentness, and futurity of events, and showing that none of them is logically possible. Thus, by a process of elimination, it is inferred that the only change possible is a change in an event's position in the *A*-Series. If time consisted only of a *B*-Series the following are the possible candidates for "change."

(i) It might be held that an event in the *B*-Series could change. Maybe the death of Queen Anne could somehow cease to be the death of Queen Anne. But this is absurd. Events can never cease to be just the sort of events they are.

Take any event—the death of Queen Anne, for example—and consider what changes can take place in its characteristics. That it is a death, that it is the death of Anne Stuart, that it has such causes, that it has such effects—every characteristic of this sort never changes.⁸

Events remain just as sweet, young, and innocent as they always were for McTaggart. By the law of identity each event must forever remain itself.

(ii) It might be contended that two events in

⁷ Surprisingly, none of McTaggart's commentators and critics have asked this basic question. The explanation for this strange oversight might be due to the ease with which we can *picture* the *A*-Series, and often this picture involves a correlation between the events of the *A*-Series and the marks on the edge of a ruler. Having such a picture of the *A*-Series we then think we understand it and forget to ask what is its generating relation. This is a good example of the danger of pictorial thinking.

⁸ H. Bergson's analysis of time, I believe, comes very close to committing this absurdity, for he denies that events are distinct from each other.

⁹ McTaggart, *op. cit.*, p. 13.

the *B*-Series could merge so as to form a new event. But such conjugal union between events is logically unthinkable for the reason given in (i): it would require that the two events which merge would cease to be themselves. If the death of Queen Anne and the death of Stalin were somehow to merge to form a new event consisting of the death of Queen Stalin then these events would no longer be the same events.

(iii) Events might be thought to have the ability to get in or out of a *B*-Series, and this could constitute the change we are seeking. But this sort of change also is logically impossible, for the *B*-Series is a sort of logically escape-proof prison. An event "can never get out of any time-series in which it once was." Nor can an event enter into a *B*-Series in which it is not a member, since positions in the *B*-Series are permanent.

(iv) Possibly change could arise from events shifting in their position in the *B*-Series? This also is logically impossible, for our concept of the generating relation of the *B*-Series—*earlier than*—is such that it is nonsensical to speak about events changing in their relations of precedence to other events or changing in the metrical features of these relations. "If *N* is ever earlier than *O* and later than *M*, it will always be, and has always been, earlier than *O* and later than *M*, since the relations of earlier and later are permanent." Events are not able to sneak up on each other.

From this McTaggart concludes that the *B*-Series cannot give us change, and therefore the *B*-Series, although necessary for the reality of time, is not sufficient. To get change and therefore time we must invoke the change in the position of an event in the *A*-Series. Such changes are the only logically possible ones. For there to be such change there must be an *A*-Series. Take away the *A*-Series and the *B*-Series ceases to be a temporal series.

A proponent of the *B*-Theory Answer would claim that McTaggart has overlooked one possible candidate for the title of "change," namely that it is things, rather than events, that change, and such change consists in a thing having a different quality at one phase of its history than it has at an earlier or later phase; e.g., a poker changes in that it is at one time hot and at a later time cool. Such qualitative changes in things can be analyzed

solely in terms of the *B*-relations of subsequence or precedence between qualitatively different states which comprise the history of a single thing. Moreover, these qualitatively different states of a single thing can be described in a tenseless language, which makes no references to the *A*-determinations, i.e., pastness, presentness, or futurity, of these states; e.g., "The poker is hot on Monday" and "The poker is not hot at times other than Monday."¹⁰ McTaggart raised the following objection to this analysis:

This makes no change in the qualities of the poker. It is always a quality of that poker that it is one which is hot on that particular Monday. And it is always a quality of that poker that it is one which is not hot at any other time. Both these qualities are true of it at any time—the time when it is hot and the time when it is cold. And therefore it seems to be erroneous to say that there is any change in the poker.¹¹

McTaggart's reply fails because of an equivocation on the word "always." He argues that because the tenseless statement "The poker is hot on Monday" is *always* true, i.e., true independently of time in the sense that the use of this sentence makes a true statement every time it is uttered, the state of affairs described by this statement must *always*, i.e., in the temporal sense of "always," be occurring or existent. Because statements describing events in a tenseless manner are *always* true it does not require that these events are sempiternal: to claim the opposite is an unwarranted addition of an eternalistic ontology to a tenseless (or token-reflexive free) logic.

McTaggart's other argument to show the dependency of the *B*-Series upon the *A*-Series fares no better than his first one. This argument, which is never stated explicitly, is that our understanding of the concept of earlier (later) than essentially involves reference to the concepts of past, present, and future. *Earlier than* is a temporal relation only because its relata have *A*-determinations. At one place he claimed that we could *define* "earlier than" in terms of the *A*-determinations of "past, present, and future."

The term *P* is earlier than the term *Q*, if it is ever past while *Q* is present, or present while *Q* is future.¹²

¹⁰ Russell's analysis of change as arising from the fact that there are propositional functions (e.g., "The poker is hot at *t*," in which '*t*' is a free variable) which are true of some, but not all moments of time, is in *The Principles of Mathematics* (London, 1903), p. 472.

¹¹ McTaggart, *op. cit.*, p. 15.

¹² *Ibid.*, p. 271.

If it is possible to define "earlier than" in terms of "past, present, and future" it will follow not only that there could be no *B-Series* without an *A-Series*, but also that the *B-Series* is reducible to the *A-Series*, since the generating relation of the *B-Series* is definable in terms of *A-determinations*.

McTaggart's attempt to define the *B-relation* "earlier than" in terms of the three *A-determinations* fails because of circularity. The *definiens* of the definition contains the statements, "*P* is past while *Q* is present" and "*P* is present while *Q* is future," which mean the same as, respectively, "*P* is past at *Q*" and "*Q* is future at *P*." The predicates "... is past at ..." and "... is future at ..." are synonyms for "... is earlier than ...," in that all three are tenseless two-place predicates which, when a non-indexical or non-token-reflexive event expressions are substituted for their blank spaces, express a timelessly true or false statement about a *B-relation* of precedence between two events.¹³ "*Is* past (present, future) at some event or date," unlike "*is* now past (present, future)," does not contain a temporal index. But, as pointed out above, an *A-determination* must contain a temporal index, for two non-simultaneous uses of the *A-determinations* of "past," "present," and "future" will determine different *A-Series* of events. The event(s) which is present in one of them will not be in the other. Therefore, McTaggart is cheating by allowing "*is* past (future) at" to count as *A-determinations*. If the copulae of the statements in the *definiens* are taken to be tensed—"P is now past while Q is now present" and "P is now present while Q is now future"—then the *definiendum* is not logically equivalent to the *definiens*; for it might be the case that *P* is earlier than *Q*, but be false that *P* or *Q* is now present, in which case the *definiendum* is true and the *definiens* false.

If we take the *A-Series* to be determined by an unqualified past, present, and future, i.e., a past, present, and future that do not admit of degrees, then it can be shown that the *B-Series* cannot be reduced to the *A-Series*, this being due to the fact that the *B-relation* "earlier than" cannot be defined in terms of an unqualified "past, present, and future." In the first place, it is fallacious to argue that *earlier than* can be analyzed in terms of *A-determinations* because there are *A-determinative* statements which imply a tenseless *B-relation* statement but no tenseless *B-relation* statement that implies an *A-determinative* statement; e.g., "*P* is present and *Q* is future" entails, but is not entailed by, "*P* is earlier than *Q*."¹⁴ For the fact that a statement containing a term *X* entails, but is not entailed by, a statement containing *Y* does not show that *Y* is analyzable in terms of *X*. For example, "All *animals* are mortal" entails, but is not entailed by, "All *men* are mortal," but this surely does not show that men are analyzable purely in terms of the concept of an animal. Thus, while it is true that there is an asymmetry in information content between a pure *B-relation* language and a pure *A-determinative* language, in that everything sayable in the former is sayable in the latter but not vice-versa, it does not establish that *B-relations* are reducible to *A-determinations*, and therefore that the *B-Series* is reducible to the *A-Series*.

Another attempt to reduce the *B-Series* to a *pure A-Series*, i.e., an *A-Series* determined by the use of an unqualified "past, present, and future," which is also unsuccessful, involves giving a "definition in use" of "earlier than" in terms of the three unqualified *A-determinations*, taken together. By a *definition in use* of a term is meant one which gives an analysis of a statement containing this term: a necessary, though possibly not sufficient, condition

¹³ It is crucial to find a language-neutral criterion for distinguishing between an *A-determinative*, or tensed, statement and a *B-relation*, or tenseless, statement. The strategy in discovering this criterion is to begin with paradigm cases of *A-determinative* statements, such as "*S* is now (was, will be) *P*," or "There now is a *P*ing of *S*," and to extrapolate from these examples a set of rules of use for the sentences employed in making these statements. These rules of use are: a present tensed sentence can be used to make a true statement only if uttered simultaneously with the event described, and so on for the other tensed sentences. Thus, an *A-determinative* statement is one which is made by the use of a sentence which has temporal restrictions placed upon its use: its use at different times may make statements having different truth-values. A *B-relation*, or tenseless, statement is one which is made through the use of a sentence which has no temporal restrictions placed upon its use: the utterance of a tenseless sentence makes a statement having the same true-value whenever it is uttered. For a detailed account of the logic of tensed and tenseless statements see my articles: "Endorsing Predictions," *The Philosophical Review*, vol. 70 (1961), pp. 376-385; "Tensed Statements," *The Philosophical Quarterly*, vol. 12 (1962), pp. 53-59; "A Reply to Smart, Mayo, and Thalberg on 'Tensed Statements'," *The Philosophical Quarterly*, vol. 13 (1963), pp. 351-356; "Is It Now Now?" *Mind*, vol. 73 (1964), pp. 97-105; "The Egocentric Particular and Token-Reflexive Analyses of Tense," *The Philosophical Review*, vol. 73 (1964), pp. 213-228; and "Existence, Tense, and Presupposition," *The Monist*, vol. 50 (1966).

¹⁴ I committed this fallacy in my paper, "Tensed Statements," *op. cit.*, and am indebted to Professor J. J. C. Smart for pointing this out in his paper, "Tensed Statements: A Comment," *The Philosophical Quarterly*, vol. 12 (1962), pp. 264-265.

for a successful analysis is that the *analysandum* and the *analysans* are logically equivalent. Let us try to analyze a tenseless *B*-relation statement in terms of a logically equivalent disjunction of *A*-determinative statements:

(I) *P* is earlier than *Q* \equiv *P* is past and *Q* is present or *P* is past and *Q* is future or *P* is present and *Q* is future

in which 'P' and 'Q' are expressions referring to an event or state of affairs. It is quite apparent that (I) is not a logical equivalence, for whereas it could not be the case that the disjunction of *A*-determinative statements in the *analysans* is true and the *B*-relation statement in the *analysandum* is false, it could be the case that the *analysandum* is true and the *analysans* is false. The reason for this, which has already been given in the above discussion of McTaggart's definition, is that if *P* and *Q* are either both past or both future, *P* being more past than *Q* or *Q* being more future than *P*, then each of the three disjuncts would be false. To give an adequate definition in use of a *B*-relation in terms of a disjunction of *A*-determinative statements we must enlarge our concept of an *A*-determination so as to include "more past" and "more future" as well as an unqualified "past," "present," and "future." With this expanded concept of an *A*-determination we can add two disjuncts to the *analysans* of (I):

(II) *P* is earlier than *Q* \equiv *P* is past and *Q* is present or *P* is past and *Q* is future or *P* is present and *Q* is future or *P* is more past than *Q* or *Q* is more future than *P*.

What is to be concluded from this discussion is that the *B*-Series cannot be reduced to a *pure A*-Series, but that it can be reduced to an *impure A*-Series, i.e., an *A*-Series which is determined not only by an unqualified past, present, and future but also by more past and more future.¹⁵ Thus the

thesis of McTaggart and the *A*-Theory Answer that the *B*-Series is reducible to the *A*-Series is correct, provided that we are speaking of an *impure A*-Series. This qualification greatly trivializes their thesis. Not only can the *B*-relation *earlier (later) than* be analyzed in terms of the *five A*-determinations, taken together, but the same is true for the relation *simultaneous with*. The latter can be classified as a *B*-relation for the following two reasons. First, the relation of *simultaneity* is needed to construct the *B*-Series. The *B*-Series includes the totality of events which make up the history of the world. The generating relation of this series—is *earlier than*—is not connected in this set of events, since some events are simultaneous with each other. Is *earlier than* related classes of simultaneous events. The relation of *simultaneity* is needed in determining the classes of events which are ordered by the relation of *earlier than* to form the *B*-Series. Therefore, the relation of *simultaneity* is needed in the construction of the *B*-Series. Second, "... is simultaneous with ..." has the same logic as "... is earlier than ..." in that both are tenseless two place predicates expressing a temporal relation which, when non-indexical event expressions are substituted for their blank spaces, make timelessly true or false statements. The following is a reduction of "is simultaneous with" to a disjunction of *A*-determinative statements:

(III) *P* is simultaneous with *Q* \equiv (*P* is present and *Q* is present) or (*P* is past and *Q* is past and it is not the case that either *P* is more past than *Q* or *Q* is more past than *P*) or (*P* is future and *Q* is future and it is not the case that either *P* is more future than *Q* or *Q* is more future than *P*).

Once again, it was necessary to use "more past (future)" in carrying out the reduction.¹⁶

In summary, an impure, but not a pure, *A*-Series is also a *B*-Series, in that a statement describing the *A*-determinations of events entails a statement de-

¹⁵ It might be pointed out that the concept of more past (future) is not a metrical concept. Even if time has no intrinsic metric (due to there being no ultimately discrete events) and we have not laid down conventional co-ordinative definitions for the metricization of time, we can say that one temporal interval is more past (future) than another due to the fact that these intervals overlap, one of them containing the other as a part of itself. For an excellent discussion of this see A. Grünbaum, *Philosophical Problems of Space and Time* (New York, 1963), chap. I.

¹⁶ The thesis of the *B*-Theory Answer that the *A*-Series can be reduced to the *B*-Series obviously is false, since an *A*-determinative statement cannot be analyzed into a tenseless *B*-relation statement, or even a disjunction of such statements. "*P* is present and *Q* is future" certainly is not logically equivalent to "*P* is earlier than *Q*, or *P* is later than *Q*, or *P* is simultaneous with *Q*," for the latter is a tautology, assuming that 'P' and 'Q' refer successfully, while the former is contingent. And even if we make the latter non-tautological by dropping one of its three disjuncts it could still be true and the former tensed statement be false. "*P* is now present" can be analyzed into "*P* is simultaneous with this (or now)," in which "this" is a demonstrative referring either to some event experienced by the speaker when he makes his utterance or to the utterance-token itself, but the latter is a tensed or *A*-determinative statement according to the criterion given in n. 13 above.

scribing tenseless *B*-relations between these events. However, a *B*-Series is not also a specific *A*-Series, in that a statement describing *B*-relations between events does not entail a statement describing the specific *A*-determinations of these events, but only a disjunction of statements describing different possible combinations of the *A*-determinations of these events. Thus, an *A*-Series already is some specific *B*-Series, but a *B*-Series is not at the same time some specific *A*-Series. This conclusion helps us to see why McTaggart's attempt to derive the *B*-Series from a conjunction of the *A*-Series with a *C*-Series, i.e., a series whose generating relation is non-temporal, is either pointless or futile.

We can now see that the *A*-Series, together with the *C*-Series is sufficient to give us time. . . . The *B*-Series . . . is not ultimate. For given a *C*-Series of permanent relations of terms, which is not in itself temporal, and therefore is not a *B*-Series, and given the further fact that the terms of this *C*-Series also form an *A*-Series, and it results that the terms of the *C*-Series become a *B*-Series, those which are placed first, in the direction from past to future, being earlier than those whose places are *further in the direction of the future*.¹⁷

If the *A*-Series referred to in this quotation is an impure *A*-Series, then it is already a *B*-Series; and therefore it is pointless, since unnecessary, to combine this *A*-Series with another series in order to derive the *B*-Series. But if the *A*-Series referred to by McTaggart is a pure *A*-Series then it can be shown that his attempt to derive the *B*-Series in the manner described is futile. Consider the problem of correlating the infinitely denumerable members of a pure *A*-Series with the members of a *C*-Series, such as the series of integers. There is no problem in correlating the present event(s) in the pure *A*-Series with the number 0 in the *C*-Series of integers. But it is impossible to correlate the past and future events of this *A*-Series respectively with the minus and plus integers because this *A*-Series contains a "blob" past and future, i.e., one which admits of no distinction between more past and less past. To make this correlation we must import a structural order into the past and future of the *A*-Series by introducing the concept of more past and more future, which, incidentally, is exactly what McTaggart does when he speaks of terms that are "*further in the direction of the future*." We

must say something like "the more past (or future) an event is in the *A*-Series the smaller (or larger) the integer with which it is correlated."

Having clarified McTaggart's ambiguous concept of the *A*-Series and shown that only the impure *A*-Series is a series, we can now answer our original question concerning the generating relation of the (impure) *A*-Series. The answer is that its generating relation is *earlier than*, which is also the generating relation of the *B*-Series. Since a *B*-Series must also be an *A*-Series, though not any specific *A*-Series, it follows that the generating relation of the *B*-Series must also generate an *A*-Series, though not any one specific *A*-Series. There is another way of answering this question. The *A*-Series is formed from a conjunction of two different series having a common member, which serves as the *terminus ad quem* of one and the *terminus ab quo* of the other. There is a series whose generating relation is *more past than*, and there is another series whose generating relation is *more future than*. These two series can be conjoined to form an (impure) *A*-Series because the *terminus ad quem* of the series of past events is the present event(s) while the *terminus ab quo* of the series of future events is also the present event(s).

Regardless of which answer we give to the question concerning the generating relation of the *A*-Series we have to invoke the concepts of more past and more future. It might be objected, and not without plausibility, that these concepts involve a tenseless *B*-relation because a statement of the form, "*x* is now more past than *y*," must be *analyzed* into a conjunction of the *B*-relation statement, "*x* is earlier than *y*" and the *A*-determinative statement, "*y* is now past." And if this is so, then our reduction of *B*-relations to *A*-determinations by logical equivalence (II) is circular, since the final two disjuncts in the *analysans* (the ones using "more past" and "more future") contain *B*-relation statements. There are two ways of countering this objection. First, it could be argued that the statement, "*x* is more past than *y*," is an *A*-determinative or tensed statement because it employs a sentence which has the same temporal restrictions placed upon its use as are placed upon the use of sentences, such as "*x* is now past," which unquestionably are used to make an *A*-determinative statement.¹⁸ The utterance of "*x* is more past than *y*" at different times may make statements having

¹⁷ McTaggart, "The Unreality of Time," *Mind*, vol. 17 (1908), p. 462. My italics. This article is reprinted in McTaggart, *Philosophical Studies* (London, 1934).

¹⁸ See n. 13 above.

different truth-values, which could not happen with different utterances of a sentence used in making a *B*-relation statement, such as "*x* is earlier than *y*." This fact, however, is not decisive in showing that this statement cannot be analyzed into a conjunction of a *B*-relation statement and an *A*-determinative statement; for the conjunction of an *A*-determinative and *B*-relation statement will always make an *A*-determinative statement, just as, analogously, the conjunction of a contingent and a necessary statement will always make a contingent statement. A more effective way to meet this charge of circularity in our reduction of *B*-relations to *A*-determinations is to claim, as we do in (II), that "*x* is more past than *y*" entails "*x* is earlier than *y*" but deny that we must analyze "*x* is more past than *y*" into the conjunction of "*x* is earlier than *y*" and "*y* is now past." I know of no argument that can be given to show that we must accept this as an analysis of "*x* is more past than *y*."

I am not very happy with this reply to the charge of circularity. It might be thought that this rather indecisive result could be avoided by offering a different reductive analysis of *B*-relations to *A*-determinations than that given in (II), e.g., instead of "more past than" we would have used "preceded," "was earlier than," or "occurred a longer time ago than," etc., but they are just as susceptible to the charge of circularity as is "more past than." We might try and recast McTaggart's question about the relation between the *B*- and the *A*-Series by using the connectives "before" and "while" instead of the relations "earlier than" (and "simultaneous with"). The former take sentences as their arguments, unlike the latter which take only event, or state of affairs, expressions as their arguments. The question then is whether we can

analyze a statement containing the connective "before" and tenseless statements for its arguments into a logically equivalent disjunction of statements containing "before" and "while" and only *A*-determinative statements for their arguments. This can be accomplished as follows:

(IV) *P* is *Q* before *R* is *S* \equiv (*P* was *Q* and *R* is *S*) or (*P* was *Q* and *R* will be *S*) or (*P* is *Q* and *R* will be *S*) or (*P* had been *Q* while *R* was *S*) or (*P* will have been *Q* while *R* will be *S*).¹⁹

Since (IV) is the analogue of (II) in connective discourse it should not surprise us to find that (IV) is haunted by the same specter of circularity as is (II). Once again the difficulty occurs in the final two disjuncts. Instead of the problem arising with "more past than," as it does with (II), it concerns the use of the compound tense "had (will have) been." One might claim that we must analyze "*P* had been *Q* while *R* was *S*" into the conjunction of the *A*-determinative statements, "*R* was *S*," with the tenseless *B*-relation statement, "*P* is *Q* before *R* is *S*," thus making the analysis circular. And once again the reply must be that there is no reason why we must accept this as an analysis. One thing is certain: by allowing, as we must, "more past (future) than" or a compound tense to count as an *A*-determination we trivialize the thesis of McTaggart and the *A*-Theory Answer that the *B*-Series is reducible to the *A*-Series. But what this indicates about our concept of time is far from trivial. It shows that our concept of time involves not only the three unanalyzable concepts of past, present, and future, but also the notion of a structural order, and it was because of the latter that we had to introduce the concepts of more past and more future.²⁰

University of Pittsburgh

¹⁹ "*S* is *P* while *R* is *T*" is logically equivalent to "*S* is *P* and *R* is *T* or *S* was *P* while *R* was *T* or *S* will be *P* while *R* will be *T*." This is the analogue in connective discourse to reduction (III).

²⁰ This paper was written while the author was engaged in research on "The Logic of Time Language" under a National Science Foundation grant. The author is indebted to Professor Edmund Gettier and Mr. Jay Rosenberg for their critical comments on a draft of the paper.

VI. POSTULATES FOR TENSE-LOGIC

A. N. PRIOR

I PROPOSE here to axiomatize six systems of tense-logic, which I shall call GH_1 , $P-$ (P minus), Pf_0 (i.e., Pf zero), Pf , O (for Occam) and O' . I shall also sketch certain further systems, namely P (for Peirce), P' , GH_2 , GH_3 , and GH_4 , and although I shall make only the most tentative moves toward the axiomatizations of these, I shall construct exact models of them. The GH systems are purely propositional tense-calculi without variables representing specific time-intervals; the others have such variables. GH_1 , $P-$, Pf_0 , and Pf are symmetrical as regards past and future; the others are not.⁰

In all the formalizations, I use the ordinary Łukasiewicz symbols for truth-functions, e.g., " $\neg a$ " for "Not a ," $Ca\beta$ for "If a then β ," $Ka\beta$ for "Both a and β ," and $Aa\beta$ for "Either a or β " (non-exclusive); and in most of them I shall also use his symbols for quantifiers, i.e., Πn for "For all n " and Σn for "For some n ." All the systems are designed for subjoining to the classical propositional calculus, with its rules of substitution and detachment extended to such new formulae as they may contain (but with substitution restricted in O and O').

I. TENSE-LOGIC WITHOUT SPECIFIC INTERVALS: THE SYSTEM GH_1

For this, we add to the classical propositional calculus the following symbols for weak and strong past and future tenses:

- Pp for "It has been the case that p "
- Hp for "It has always been the case that p "
- Fp for "It will be the case that p "
- Gp for "It will always be the case that p ."

G and H are primitive, and the other two defined as follows:

Df. $F : F = NGN$

Df. $P : P = NHN$

The system has the following special rules:

RG: If $\vdash a$ then $\vdash Ga$

MI (the Mirror Image rule): In any thesis we may replace P by F , G by H , and vice versa, throughout.

(For example, since $CpGpp$ is a thesis, so is $CpHFp$.) The axioms are as follows:

- A1. $CGCpqCGpGq$
- A2. $CGpNGNp (= CGpFp)$
- A3. $CGpGGp$
- A4. $CGGpGp$
- A5. $CGCpqCGCpGqCGCFpqCFpGq$
- A6. $CNHNp (= CPGpp)$
- A7. $CpCGpCHpGHp$

These axioms are consistent, since if we let $Gp = Hp = p$ they become laws of the propositional calculus. (This was shown by T. Smiley.) A4 is independent of the rest, since it, and it alone, depends on the assumption that time is continuous. If there were "next moments," GGp could be true if p only began to be true the moment after the next (remaining true thereafter), whereas Gp can only be true if p is already true by the next moment (after the present one). I have no proofs of independence for the other axioms, except that in the absence of the mirror-image rule we have all but A3 if we read Hp as p and G as the necessity operator L of Feys' modal system T , plus Geach's special axiom for the system $S4.3$, $ALCLpqLCLqp$; all but A5 if we read Hp as p and G as the L of $S4$; all but A6 if we read both H and G as the L of $S4.3$; and all but A7 if we read H as P and P as H .¹ But there may be ways of using MI to prove some of these from the others. A3, A4, and A5 are, however, definitely independent (by the above

⁰ Since the following was written, A7 in the system GH_1 has been proved from RG, MI, A1, A5, and A6 by E. J. Lemmon, and independence proofs found for A1 through A6 by D. Berg and I. M. Hacking. It has also been observed, by R. Montague and N. Cocchiarella, that a stronger system than GH_1 results if the series of moments is assumed to be not merely dense, like the rationals, but strictly continuous, like the reals. "Rational" should therefore be put for "real" throughout Section II.

¹ For the modal systems here referred to, see A. N. Prior, "Tense-Logic and the Continuity of Time," *Studia Logica*, vol. 1 (1962), pp. 113-151. I shall refer to this in future notes as TLCT.

proofs) in the purely future-tense sub-set of the postulates, i.e., Df.F, RG, and A1-A5. Whether these suffice for the purely future-tense portion of the system, I do not know, but I suspect that they do.

We may begin the development of the system with the following simple proofs:

T1. $GCCpGqNqNp$	[p.c. RG]
T2. $CGCpGqNqNp$	[T1, A1]
T3. $CGCpGqNqGp$	[T2, A1, p.c.]
T4. $CGCpGqNqGpNq$	[T3, p.c.]
T5. $CGCpGqCFpFq$	[T4, Df. F]
T6. $CHCpGqCHpHq$	[A1, MI]
T7. $CHCpGqCPpPq$	[T5, MI]

And if we write " $\neg a$ " for the mirror image of a , we can establish a mirror image of RG as follows:

If $\vdash a$ then $\vdash \neg a$	[MI]
If $\vdash \neg a$ then $\vdash G(\neg a)$	[RG]
If $\vdash G(\neg a)$ then $\vdash \neg G(\neg a)$	[MI]
But the mirror image of $G(\neg a)$ is $H(a)$.	
Hence, if $\vdash a$ then $\vdash Ha$.	

This last, which we may call RH, can be used with T6 and T7 in the same way as RG can be used with A1 and T5, to establish the extensionality principles

If $\vdash Ca\beta$, then $\vdash CGaG\beta$, $\vdash CFaF\beta$, $\vdash CHaH\beta$ and $\vdash CPaP\beta$.

These enable us to interchange logically equivalent formulae, e.g., p and NNp , in all tensed contexts. For example, we may carry out the following derivation from A5:

T8. $CKKGCpGqGqGCFpGqCFpGq$	[A5, p.c.]
T9. $CNCFpGqNKKGCpGqGqGCFpGq$	[T8, p.c.]
T10. $CNCFpGqNqNKKGCpNqGqGqGCFpNq$	[T9, q/Nq]
T11. $CNCFpNFqNKKNFNCpNqNFNCpNFqNFNCpNFq$	[T10, $GN = NF$, $G = NFN$]
T12. $CKFpFqNKKNFpFqNFKpFqNFKFpFq$	[T11, $NCpNq = KpFq$]
T13. $CKFpFqAAFpFqFKpFqFKFpFq$	[T12, de Morgan]

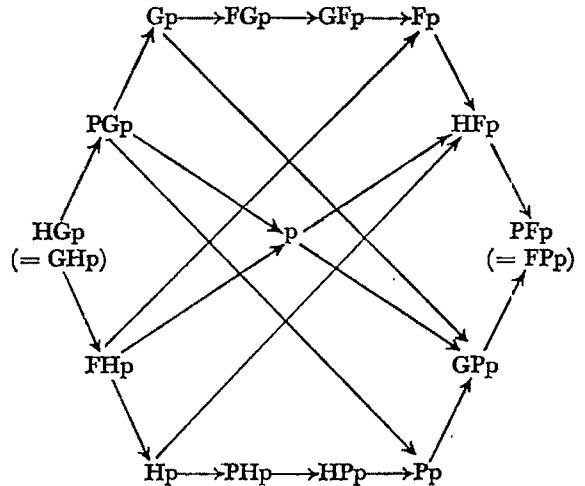
This last thesis states that if each of p and q is going to be the case ($KFpFq$) then either they will be true together ($FKpFq$), or p will occur and q after it ($FKpFq$), or q will occur and p after it

($FKpFq = FKpFq$). It expresses the *linear* character of time much more succinctly than A5 itself, but A5 lends itself more directly to other symbolic manipulations. For example, if we put Gp for p and Fp for q in A5 we obtain

$$CGCGpFpCGCGpGFpCGCFpFpCFpGFp$$

Here the antecedents $GCGpFp$, $GCGpGFp$, and $CGCFpFp$ may all be detached for various reasons. (We have the first from A2 and RG; the second from A3, S2, and G -extensionality; and the third via $CFpGFp$, from A2 and F -extensionality, and $CFpGFp$, from A3 by transposition.) We can thus assert the final consequent $CFpGFp$, "If it will be that it will always be that p , then it will always be that it will be that p ."

A5 cannot be replaced (at all events in the pure FG portion of the system) by what we have just proved from it, $CFpGFp$; as this does not fully express the linear character of time, but would be equally true if time forked at places, provided that the branches always came together again sooner or later.² But even in the weakened system we would have enough to prove all the equivalences required for what may be called Hamblin's Theorem; namely, that if by a "tense" we mean some sequence (possibly null) of prefixes drawn from P , F , H , G , there are precisely 15 non-equivalent tenses, none of them with more than two prefixes, with their order of strength as in the following diagram:³



² Cf. the last section in A. N. Prior, "K1, K2 and Related Modal Systems," *Notre Dame Journal of Formal Logic*, vol. 5 (1964), pp. 299-304.

³ The method of proving this theorem is sketched in TLCT. The implications $CGpGp$, $CPGpPp$, $CFHpFp$, and $CHpHfPp$ were not noticed by Hamblin until 1965.

If, following Diodorus, we define Lp , i.e., "Necessarily p ," as $KpGp$, "It is and always will be that p ," the L-fragment of GH_1 becomes the modal system $S_4.3$.⁴ Hintikka, it may be noted, has axiomatized $S_4.3$ by adding to S_4 the axiom $CKMp-MqAMKpMqMKMp$, the proof of which from $Df.L$ and T_{13} is obvious. If Lp is defined as $KKpHpGp$, "It is and always has been and always will be that p ," the L-fragment becomes S_5 .

We can also define in GH_1 certain functions on which Miss Anscombe has recently written,⁵ namely Tpq , "It was the case that p and then it was the case that q "; Bpq , "It was the case that p before it was the case that q "; and Rp , "It has been the case repeatedly that p ." $Tpq = PKPpq$, i.e.,

It has been the case that at once

(1) it has been the case that p , and

(2) (it is the case that) q .

$Bpq = TKpNqq$ ("It was the case that p -but-not- q and then it was the case that q ."). $Rp = TTPNpp$ (" p and then not p and then p "). We can prove such theses as $CTTpqTTpq$ (if we had p and then q and then r , we had p and then q), $CTTpqTTqr$, $CTTpqTTpr$, and $CKpFqFTpq$, i.e., if p is now true and q will be, then (sooner or later) we will have had p -and-then- q .

II. TENSE-LOGIC WITH SPECIFIC INTERVALS:

THE SYSTEMS P—, PFO, AND PF

In these systems we employ the forms Pnp and Fnp for "It was the case the interval n ago that p " and "It will be the case the interval n hence that p ," and distinguish the weaker and stronger tenses by existentially and universally quantifying over intervals. It is convenient to regard the variables n, m , etc., as standing for real numbers measuring intervals. In the system P—, the entire range of real numbers—positive, negative, and zero—is used for this purpose, and an interval hence is defined rather artificially as a negative interval ago. In PFO, this artificiality is dropped; P and F are independent primitives, and n, m , etc., stand for non-negative real numbers only. Zero, however, is retained in PFO, both Pop and Fop being equated with the plain p ("It is the case that p "), i.e., the present is treated as a limiting case of the future

and the past. In PF even this is dropped: n, m , etc., stand only for positive real numbers, Pop and Fop being meaningless. In each system, we may substitute for m, n , etc., throughout a thesis, any expressions denoting real numbers from the range involved, e.g., $m + n, m - n$ (subject in PFO to the proviso that $m \geq n$, and in PF to the proviso that $m > n$) or 0 (except in PF). Further, at any point in a thesis we may substitute for such a variable any expression denoting the same real number, e.g., for m we may substitute $n + m - n$.

These systems are for subjunctive not only to the propositional calculus, but also to any normal basis for quantification theory, e.g., the Lukasiewicz rules II_1 and II_2 with the definition of Σn as $NIInN$. The only variables subject to quantification, however, are m, n , etc.

Turning now to P—, this has (in addition to propositional calculus and quantification theory) simply the definition

Df. F: $Fnp = P(-n)p$,

the rule

RP: If $\vdash a$ then $\vdash Pna$,

and the axioms

Po: $CPopp$

PN₁: $CPnNpNPn$

PN₂: $CNPnpPnNp$

PC: $CPnCpqCPnpPnq$

PP: $CPmPnpP(m+n)p$

PII: $CIIInPmPnpPmIIPnp$.

The converses of Po, PC, PP, and PII are all provable.⁶ From RP and PC we have the rule:

If $\vdash Ca\beta$ then $\vdash CPnaPn\beta$,

and through this we have the broader rule that if α and β are logically equivalent—as $PnNp$ and $NPNp$, in particular, are by PN₁ and PN₂—they are interchangeable in any formula of the system. Through this we have, e.g., the following:

T ₁ : $CPnKpqPnp$	[from $CKpq$]
T ₂ : $CPnKpqPnq$	[from $CKpq$]
T ₃ : $CPnKpqKPnpPnq$	[T ₁ , T ₂]

⁴ This result follows from R. A. Bull's "An Algebraic Study of Diodorean Modal Systems," *The Journal of Symbolic Logic*, vol. 30 (1965), pp. 58-64.

⁵ G. E. M. Anscombe, "Before and After," *The Philosophical Review*, vol. 73 (1964), pp. 3-24.

⁶ These results I have derived essentially from N. Rescher, "On the Logic of Chronological Propositions," *Mind*, vol. 75 (1966), where some of them are credited to R. Meyer.

- T4: $CNKpnpPnqNPnKp$ [T3, p.c.]
 T5: $CNKpnpPnNqNPnKpNq$ [T4]
 T6: $CNKpnpNPnqPnNKpNq$ [T5; $PnN = NPn$]
 T7: $CCpnpPnqPnCp$ [T6; $NKpNq = Cp$]

T7 is the converse of PC. Another important theorem in P— is the conclusion of the following deduction:

- T8: $CIIInPnpPmp$ [Quantification theory]
 T9: $CIIInPnpP(n - m)p$ [T8]
 T10: $CPmIIInPnpPmP(n - m)p$ [T9, RP, PC]
 T11: $CPmIIInPnpP(m + n - m)p$ [T10, PP]
 T12: $CPmIIInPnpPnp$ [T11; $m + n - m = n$]
 T13: $CPmIIInPnpIIInPnp$ [T12; 2n]

This asserts in effect that if it is true at any time that p is true at all times, then it is true now that p is true at all times. It means that the truth-value of $IIInPnp$ is itself time-independent.⁷ In PFO and PF, we may note, it does not mean quite this, but means rather that if it has been the case that p was true at all *past* times, it is now the case that p was true at all past times. This, of course, is not true, and T13 is not provable in PFO and PF. The passage from T8 to T9 is there blocked by the consideration that $n - m$ is only substitutable for m if m is not greater than n .

If we define the necessity-functor L in P— as $IIInPn$, the resulting L-fragment is the modal system S5.

Turning now to the system PFO, in which n, m , etc., range not over all real numbers but only over non-negative ones, we must here drop Df.F, and add to the remaining postulates of P— the following rules:

- MI: In any thesis we may replace P by F , and vice versa, throughout.
 Rmn: If any formula is a thesis under the proviso that $m > n$, and also under the proviso that $n > m$, it is a thesis *simpliciter*;

and the following axioms:

- PF1: $CPmFnpP(m - n)p$, for $m > n$
 PF2: $CPmFnpF(n - m)p$, for $n > m$
 PFII: $CIIInPmFnpPmIIInFnp$

If in this system we define L as $IIInFn$, the resulting L-fragment is the modal system S4.3.

In the next system, PF, where n, m , etc., range over positive real numbers only, we must drop not

only Df.F but also the axiom Po as meaningless, and re-state Rmn as follows:

- Rmn: If any formula is a thesis under the proviso that $m > n$, and also under the proviso that $n > m$, and also under the proviso that $m = n$, it is a thesis *simpliciter*,

and replace PF1 and PF2 by the following:

- PF1: $CPmFnpP(m - n)p$, for $m > n$
 PF2: $CPmFnpF(n - m)p$, for $n > m$
 PF3: $CPnFnp$

(PF3 in effect replaces Po, much as MI replaces Df.F of P—.)

If in PF we define G as $IIInFn$ and H as $IIInPn$, its GH fragment will be the system GH1. And as an example of a proof in PF we may give a proof within it of $CpCHpCGpHGP$, i.e., the mirror image of A7 of GH1. The theses T1, T6, and T9 are the converses of PF1, PF2, and PF3, provable from these by the method used to prove the converse of PC at the beginning of this section:

- T1. $CP(m - n)pPmFnp$, for $m > n$
 T2. $CIIInPnpP(m - n)p$, for $m > n$ [Quantification theory]
 T3. $CIIInPnpPmFnp$, for $m > n$ [T1, T2]
 T4. $CHpPmFnp$, for $m > n$ [T3, Df.H]
 T5. $CpCHpCGpPmFnp$, for $m > n$ [T4; p.c.]
 T6. $CF(n - m)pPmFnp$, for $n > m$
 T7. $CGpPmFnp$, for $n > m$ [from T6, as T4 from T1]
 T8. $CpCHpCGpPmFnp$, for $n > m$ [T7; p.c.]
 T9. $CpPnFnp$
 T10. $CpPmFnp$, for $m = n$ [T9]
 T11. $CpCHpCGpPmFnp$, for $m = n$ [T10; p.c.]
 T12. $CpCHpCGpPmFnp$ [T5, T8, T11, Rmn]
 T13. $CpCHpCGpIIInPmFnp$ [T12, II2]
 T14. $CpCHpCGpPmIIInFnp$ [T13, PFII]
 T15. $CpCHpCGpIIInPmIIInFnp$ [T14, II2]
 T16. $CpCHpCGpHGP$ [T15; Df.G; Df.H]

Note that we cannot pass directly from each of T1, T6, and T9 to $CP(m - n)pCF(n - m)pCpPnFnp$, by $CCpsCpCqCrs$, $CCqsCpCqCrs$ and $CCrsCpCqCrs$ respectively, and so to the long formula without any proviso. For on each proviso, the long formula will be, for one reason or another, ill-formed. For example, since T1 is only true for $m > n$, we cannot, on this proviso substitute $F(n - m)p$ for q in $CCpsCpCqCrs$. On this proviso $n - m$ will not be a positive real number, and $F(n - m)p$ therefore not a proposition of the system.

⁷ This is used as an axiom in the paper by N. Rescher mentioned in the preceding note.

III. POWER AND THE PAST: THE OCCAMIST SYSTEMS O AND O'

In the systems discussed in this and the next sections we abandon the symmetry between past and future which characterizes all of the preceding systems and attempt to embody the idea that there are some propositions which are still within our power to make true or false, and others which are not, and that in general the past is less within our power than the future is. In the systems O and O', there are nevertheless *some* past-tense propositions which are within our power to make true or false, namely those which are (as Occam put it) equivalent to future-tense propositions. For example, the proposition "It was the case yesterday that we would be smoking two days later" is equivalent to "We will be smoking tomorrow," and is as much within our power to make true or false as the latter is. In the systems P and P' (discussed in the next section) equivalences such as this one do not hold, and the making true or false of past-tense propositions is *totally* outside our power; while future-tense propositions are considered as being simply false until it is beyond our power to make them so.⁸

In the "Occamist" systems O and O' we formalize these conceptions by introducing a necessity-functor *L* such that *Lp* is true if and only if it is beyond our power to make *p* false; and substitution for propositional variables is restricted to formulae—that we call A-formulae—which represent propositions which are beyond our power to make true or false. Formulae representing other propositions are, however, present in the system. Formulae in general are recursively defined as follows:

- (1) Propositional variables are formulae.
- (2) If *a* and *β* are both formulae, so are *Na*, *Caβ* (*Kaβ*, etc.), *IIa*, *Σa*, *Pna*, *Fna* and *La*.
- (3) There are no others.

A-formulae constitute a sub-class of these, defined as follows:

- (1) Propositional variables are A-formulae.
- (2) If *a* and *β* are both A-formulae, so are *Na*, *Caβ* (*Kaβ*, etc.), *IIa*, *Σa*, and *Pna*.
- (3) If *a* is any formula, *La* is an A-formula.
- (4) There are no others.

Pnp, for example, is an A-formula by clause 2 of the definition, but *Fnp* is not; nor, consequently, is *PmFnp*. *LFnp*, however, is an A-formula, and so in consequence is *PmLFnp*, though *FmLFnp* is not.

The tense-logical basis of O is the system PFO, and of O' the system PF, but because of the restrictions on substitution the axioms of these systems are replaced by axiom-schemata. For example, since *CPnNpNPnp* is an axiom of PFO, all formulae which result from putting a formula in place of *a* in the axiom-schema *CPnNaNPna* are axioms of O. (For example, *CPnNFnpNPnFnp* is an axiom of O.) Also, the mirror-image rule of PFO and PF is restricted in O and O' to theses not containing *L*. In each case, we form the Occamist system, modified as just indicated, the following rule and axiom-schemata for *L*:

- RL: If $\vdash a$ then $\vdash La$
 L1: *CLaa*
 L2: *CLCaβCLaLβ*
 L3: *CNLaLNLa*
 LF: *CLFnaFnLa*
 LFI: *CIIInFmLFnaFmIIInLFna*
 LPII: *CIIInFmLPnaFmIIInLPna*

and the axiom

- LA: *CpLp*

RL, L1, L2, and L3 give us for this undefined *L* the modal system S5. The restriction on substitution for propositional variables means that from LA we can obtain, for example, *CPnpLPnp*, but not *CFnpLFnp* or *CPmFnpLPmFnp*.

In the metatheory of the Occamist systems, we may define an O or an O' model as a line (without beginning or end) which breaks up into branches as it moves from left to right (this being interpretable as a movement from past to future), so that from any point on it there is only one route to the left (backward into the past) but a number of alternative routes to the right (forward into the future). In each O or O' model there is a single *designated route* from left to right, taking one direction only at each fork. This represents the actual course of events. And in each model, formulae are assigned truth-values (truth or falsehood) at each point on the line, in accordance with the following prescriptions:

⁸ For the detailed philosophical motivations of these systems, and the contributions of Occam and Peirce to the formulation of the alternatives which they enshrine, see A. N. Prior, "The Formalities of Omniscience," *Philosophy*, vol. 37 (1962), pp. 114-129.

(1) Each propositional variable is arbitrarily assigned a truth-value at each point, this being its *actual* assignment and its only *prima facie* assignment in the model.

(2) A *prima facie* assignment to Fna at a point x gives to it the value assigned to a at some connected point the distance n to the right of x . (If the line forks within this distance, there will therefore be a number of *prima facie* assignments to Fna at x .)

(3) The *actual* assignment to Fna at x gives it the value assigned to a at the distance n to the right of x along the designated line. If x is not on the designated line, Fna has no actual assignment there.

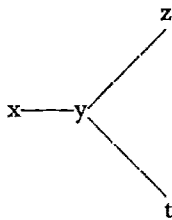
(4) The actual or *prima facie* assignment to Pna at x is the actual or *prima facie* assignment to a at the distance n to the left of x , along the line on which x lies; *except* that if a has several *prima facie* assignments at the latter point, (call it y), and none of them depends on the value assigned to a simple variable at points to the right of x on lines through x , we count (in evaluating Pna) only that *prima facie* assignment to a which uses the line yx .

(5) The actual and only *prima facie* assignment to La at x gives it truth if a is given truth in all its *prima facie* assignments at x ; otherwise falsehood.

(6) Truth-functions and quantifications are dealt with in the usual way.

A formula is a thesis of O or of O' if and only if it is true in all assignments at all points in all O or O' models. The difference between the two types of model lies simply in the values of n that can appear in Pna and Fna (zero being such a value in O but not in O').

Consider, for instance, the following simple portion of a model:



where $xy = m$, $yz = n$, the route xyz is designated, and the proposition p is true at x , y , and z and false at t . Because p is true at z , the actual value of $F(m+n)p$ at x is truth, and that of $PmF(m+n)p$ at y is therefore truth also. But because p is false at t , the *prima facie* value of $F(m+n)p$ at x when we use the route xyt will be falsehood, and the same *prima facie* assignment to $PmF(m+n)p$ will give it falsehood at y , and consequently the *actual* assign-

ment to $LPmF(m+n)p$ at y will be falsehood. Hence $CPmF(m+n)pLPmF(m+n)p$ is false at y , and is not a thesis of O or of O' ; nor is the schema $CPnaLPna$, of which it would be a particular case, a theorem-schema of O or of O' . (In the Occamist systems we have *some* power over the past.)

Here, it will be noticed, the *prima facie* assignments to $F(m+n)p$ at x do depend on the values assigned to p at points beyond y , so we have to make assignments to $PmF(m+n)p$ at y corresponding to the alternative assignments to $F(m+n)p$. On the other hand, if we were evaluating, say, $PnFnp$ at z , no assignments to Fnp at y depend on assignments to p at points past z along yz , so the only assignment to $PnFnp$ is that based on the assignment to Fnp at y which uses the line yz , making Fnp true at y and $PnFnp$ true at z . This being the *only* assignment to $PnFnp$ at z , $LPnFnp$ will be true there also; and $CPnFnpLPnFnp$ is in fact provable in O and O' , though not by simple substitutions in $CPnpLPnp$. (Since a variable p stands only for A-formulae, i.e., for formulae representing proportions that have nothing *seriously* future about them, its equivalent $PnFnp$ has nothing *seriously* future about it either, and by the time it can be said it is beyond our power to make it true or false.) On the other hand, since $LFnp$ is false at y , $PnLFnp$ is false at z , so we do not have $CLPnFnpPnLFnp$, or, in general, $CLPnaPnLa$ (the minor image of LF), though its converse is easily provable.

IV. POWER AND THE PAST: THE PEIRCEAN SYSTEMS P AND P'

The Peircean systems P and P' contain all and only those theses of O and O' which have no occurrences of F except as immediately preceded by an L , these theses being then modified by the deletion of L throughout. For example, $CpLFnFnp$ is a thesis both of O and of O' (derivable syllogistically from $CpLp$ and $CLpLFnFnp$, the latter being derivable from $CpFnFnp$ by RL and L2); hence $CpFnFnp$ is a thesis both of P and P'. On the other hand, $CpPnLFnp$ is not a thesis of O or of O' , hence $CpPnFnp$ is not a thesis of P or of P'. In the Peircean systems, in other words, it is a law that if anything is true now it will be true any interval n hence that it was true the interval n before that, but it is not a law that if anything is true now it was true an interval n ago that it would be true the interval n after that (it may, that time ago, have been preventable, and we are only entitled to say " p will be so," in the Peircean sense, when it is already

determined that it will be). Again $CLFnNpNLFnp$ is easily provable in O and O', but $CNLFnpLFnNp$ is not, so that $CFnNpNFnp$ is a thesis of P and of P', but not $CNFnpFnNp$.

Although we arrive at P and P' via O and O', the deletion of L makes these systems, in form, sub-systems of the tense-logical PFO and PF. Their symbols are the symbols of PFO and PF (P-variables, interval-symbols, truth-functors, quantifiers, and beyond those, just P and F). Also because all the formulae which occur in them are ones which in O and O' would be A-formulae, we can drop the device of axiom-schemata and revert to a finite set of axioms with an unrestricted substitution rule (or one restricted only by the usual provisos in the presence of quantifiers which also apply in PFO and PF). But certain laws of PFO and PF are absent from P and P'. Most notably, they lack the mirror-image of PN2 ($CNFnpFnNp$; call it FN2). This means in the first place that the mirror-image rule must go, so that it will be necessary at least in a first draft of an axiomatization of P and P' to lay down separately those mirror-images-of-axioms which we do have, e.g., those of RP, PN1, PC, PP, PII and the PF's. The absence of FN2 also means that we cannot use it, as we can in PFO and PF, to prove the converses of certain axioms and of their mirror-images (e.g., $CCFnPFnqFnCpq$, the converse of the mirror-image of PC), so that these also, where the system contains them, may also have to be laid down separately. The converses of the PF's, incidentally (e.g., $CP(n - m)p PnFmp$, for $n > m$) are not Peircean theses.

Although the axiomatization of the Peircean systems presents many problems, the concept of a P or P' model can be quite precisely defined. Basically a P or a P' model is the same as an O or an O' one, but with no designated route, and with a slightly modified method of assigning truth-values. There is no designated route because from the Peircean point of view there is no such thing as the actual total course of events. There is indeed an actual past—the one route that leads back from whatever point represents the present moment—but there is no actual future until we reach it and it is future no longer. Or rather, the only future that is yet actual is that much of the future which is already beyond prevention because of the momentum of the present and the past ("present in its causes," as the Thomists say).

The truth-value assignments in a Peircean model are as follows:

- (1) and (2) Assignments to propositional variables, and *prima facie* assignments to Fna , as in an Occamist model.
- (3) The actual assignment to Fna at x gives it truth if all its *prima facie* assignments at x do so; otherwise falsehood.
- (4) The actual and only *prima facie* assignment to Pna at x gives it the value actually assigned to a at the distance n to the left of x (on the connected line).
- (5) Truth-functions and quantifications as usual.

For example, if we again use the model xyz^t , where the line xy forks at y into yz and yt , where $xy = m$ and $yz = yt = n$, and p is true at x , y , and z but false at t , the falsehood of p at t gives $F(m + n)p$ one *prima facie* assignment of falsehood at x . Therefore an actual assignment of falsehood at x , and so an assignment of falsehood to $PmF(m + n)p$ at y . Also, although p is true at z , $PnFnp$ is not true there, since Fnp is false at y . Hence $CpPnFnp$ (the converse of PF3) is not a Peircean thesis.

To the Occamist, Peircean tense-logic is incomplete; it is simply a fragment of his own system—a fragment in which contingently true predictions are, perversely, inexpressible. But Peircean tenses can also be explicated without reference to Occamist ones, in terms of Peircean models. And to the Peircean, the Occamist seems to treat the future as if it were merely something that *has been* future—as if we were looking back on it from the end of time. For the Peircean can give a sense in his own language to *past* Occamist futures, provided they are *far enough* past. To be precise, he can give a sense to the Occamist "It was to be" $PmFnp$, and even to "It was contingently to be," $KPmFnpNPmLFnp$, provided that $m \geq n$. The former is simply, in the language of the system P, $P(m - n)p$; the latter, $KP(m - n)pNPmFnp$. Using "was" for the Peircean past, "was" for the Occamist, and similarly with "will," "It was the case the day before yesterday that it would be the case a day later that p " is just "It was the case yesterday that p " (at least, this works if p is not itself too future-infected). And "It was, but was not necessarily, the case the day before yesterday that it would be the case a day later that p " is just "It was the case yesterday that p ; but the following was not the case the day before yesterday: It *will* be the case a day hence that p ." Again, the meta-linguistic "Your statement of yesterday, 'It will be the case tomorrow that p ', was true," goes into Peircean as "It was the case yesterday that you were saying, 'It will be the case tomorrow that p ';

and today it is the case that p .”⁹ What cannot be said in Peircean is the Occamist’s “It is to be” (where this does not mean “It is bound to be”), i.e., his Fnp , or his $PmFnp$ where $n > m$. However, even of this it can be said in a Peircean metalanguage, “If an Occamist now says ‘It will be the case tomorrow that p ,’ it WILL (unpreventably will) be the case tomorrow that either his statement was true or it was false” (the dependent “was” being translatable as in the preceding example). This is just a case of the theorem (of P and P') $CpFnAKPnpqKPnpNq$, a consequence of $CpFnPnp$. In brief, apart from a few special cases, the Occamist tenses are not translatable into the Peircean; but the Peircean can describe their use in his own language, and could teach other Peirceans to talk Occamist. He cannot, indeed, say under what general conditions an Occamist future-tense proposition is true, as opposed to the conditions under which it *was* true (an Occamist can say that “It will be the case tomorrow that p ” is true if and only if it will be the case that p ; but this is untranslatable into Peircean). Nor, where an issue is still undecided, can the Peircean know, or tell other Peirceans, whether it is or is not correct to make a particular Occamist future-tense statement, but can only know *afterward* whether it *was* correct to make it; but at this point the Occamist himself is in exactly the same position.

V. PEIRCEAN GH SYSTEMS

If, in the system P , we define Lp in the Diodorean fashion as $IIInFnp$, we no longer obtain as our L -fragment the modal system $S4.3$ (as we do if we use this definition in the system PFO), but only the system $S4$. If Mp (“Possibly p ”) is not defined in the usual way as $NLNp$ (i.e., in P , $NIInFnNp$), but defined independently of L in P as $\Sigma nFnp$ (i.e., $NIInNFnp$), we obtain a system contained in $S4.3$ but independent of $S4$. It has such theses as $CMLpLMp$, which is not in $S4$, but lacks such $S4$ theses as $CNMpLNp$.

Similarly, if in the system P' we define G as $IIInFn$ and H as $IIInPn$, we do not obtain, as the system’s GH -fragment, our original system $GH1$, but something weaker (call it $GH2$), which lacks, for instance, the law $CFGpGFp$, i.e., $CNGNGpGNGNp$. If, however, we define the monadic F independently

of G as ΣnFn , and P as ΣnPn , we obtain yet another $PHFG$ -system (call it $GH3$) in which $CFGpGFp$ is indeed provable, but not, for example, $CpHFp$ (although its mirror image $CpGpp$ holds). $A7$, $CpCGpCHpGHp$, is provable in both $GH2$ and $GH3$, but its mirror-image, $CpCHpCGpHGp$, in neither. If p is true now, and has always been true, and it is (now-unpreventably) always going to be true, it does not follow that it has always been (unpreventably) always going to be true; though it does follow that it is (now-unpreventably) always going to be true that it always has been true.

One defect of the present formalism for Peircean tense-logic ought to be mentioned. It has no means of separating quantification over alternative future routes from quantification over intervals, and therefore no means of saying at least one thing that a Peircean might very well want to say. If, in the system P' , we define Gp as $IIInFnp$, this means that p is true throughout the whole length of all alternative futures. Fp , defined as $NLNp$, means that p is true at some future time in some alternative future; and $\bar{F}p$, defined as $\Sigma nFnp$, means that there is some interval n such that p will be true that interval n hence in all alternative futures. This last is a rather recondite way of using the plain “will,” as opposed both to just “possibly will” and to “will always”; it would be more natural to mean by it “In every alternative future there is some interval n , not necessarily the same in each alternative, such that p will be the case that interval hence in that alternative.” But this statement eludes the machinery of P and P' .¹⁰

We can, however, obtain the kind of F we want by digging back again into Occamist logic. For all the systems, Hp is defined both in O' and in P' as $IIInPnp$ and Pp as either $NHNp$ (i.e., $NIInPnNp$) or $\Sigma nPnp$ (i.e., $NIInPnNp$), these alternatives being equivalent. But since the Peircean Fn is the Occamist LFn , the definitions of Gp in O' become $IIInFnp$ for $GH1$, but $IIInLFnp$ (which is equivalent to $LIIInFnp$) for $GH2$ and $GH3$. The definitions of $\bar{F}p$ in O' become:

- (i) For $GH1$, $NGNp$ (i.e. $NIInFnNp$), which is equivalent in O' to $\Sigma nFnp$ (i.e., $NIInNFnp$).
- (ii) For $GH2$, $NGNp$, which now means $NIInLFnp$, which is equivalent in O' to $M\Sigma nFnp$, or $\Sigma nMFnp$.
- (iii) For $GH3$, $\Sigma nLFnp$ (i.e., $NIInNLFnp$).

⁹ Cf. G. Ryle, “It Was to Be,” in *Dilemmas* (Cambridge, 1962).

¹⁰ This is the defect alluded to on p. 93 of my *Time and Modality*; I do not think the solution there offered is very satisfactory. It is based on the idea that if it is true that, *whatever happens*, it will sooner or later come to pass that p , there is always a time-limit set upon when it will do so; but this is a dubious postulate.

And to get the monadic F sought in the last paragraph, we replace the last definition by $L\Sigma Fn p$. This is different from the last, and is not definable in P' because the Occamist Fn only appears in P' with an L immediately preceding it. The system in which Fp is, in terms of O' , $L\Sigma Fn p$, we may call GH_4 . Tabulating these definitions in O' for quick comparison, we have

	Gp	Fp
For GH_1 :	$\Pi nFn p$	$\Sigma nFn p$
For GH_2 :	$\Pi nLFn p$	$\Sigma nMFn p$
For GH_3 :	$\Pi nLFn p$	$\Sigma nLFn p$
For GH_4 :	$L\Pi nFn p$	$L\Sigma nFn p$

The system GH_4 resembles GH_3 in containing $CFGpGHp$ and in lacking $CpHFp$.

These definitions need not be regarded as having any more serious significance than that of explaining our meaning to Occamists. We could equally well define the different GH systems simply as

those containing as theses all formulae verified in all GH models of the relevant sort. GH models are branching lines of the sort used in O and P models, with appropriate truth-value assignments; for example, the assignments in a GH_4 model are as follows:

- (1) Each propositional variable has an arbitrary assignment of truth or falsehood at each point on the line.
- (2) Ga is assigned truth at x if a is assigned truth at every point to the right of x on every line connected to x ; otherwise falsehood.
- (3) Fa is assigned truth at x if a is assigned truth at some point to the right of x on every line connected to x ; otherwise falsehood.
- (4) Ha is assigned truth at x if a is assigned truth at all connected points to the left of x .
- (5) Pa is assigned truth at x if a is assigned truth at some connected point to the left of x .
- (6) Truth-functions as usual.

University of Manchester

VII. THE GROUNDS OF UNIVERSALITY IN ARISTOTLE

JOSEPH OWENS

I

AS a philosophical term, "universal" (*katholou*) seems clearly enough to originate with Aristotle. It does not occur in a technical sense prior to the Aristotelian treatises;¹ and is used regularly throughout their pages. Its precise meaning, moreover, stems definitely from Aristotelian philosophy, though with a deep background in Socrates and Plato. A glance at the background should help make clear on the one hand the continuity of the trend that culminated in the Aristotelian universal, and on the other hand the distinctiveness and the originality of the conception of the universal in the Stagirite's own thought. A sufficient grasp of what the universal meant for him, and how he saw it functioning in scientific knowledge, is an obvious preliminary requirement for an investigation into its grounds.

With Socrates the notion of what something is, in contrast to the particular instances in which it is found, had been stressed consistently against the teaching of the Sophists. In Plato this notion was carried over into the Idea, separate from and participated in by the particular sensible things. For both Socrates and Plato the all-pervading concern with the topic arose from the need to base human thought and human conduct upon scientific knowledge, rather than upon the traditional doxastic conceptions inculcated so successfully by the Sophists and by the flourishing school of

Isocrates.² In the struggle for scientific knowledge, Aristotle was continuing the trend inaugurated by Socrates and Plato. As with his two predecessors, his interest lay in grounding human knowledge and activity upon the permanently stable, instead of locating it in a collection of moment-to-moment opinions upon an ever-changing flux. In the Aristotelian noetic, however, only particular sensible things could be the original objects of human cognition.³ They alone provided the data for all further knowledge. Of their nature they were changeable. Yet within them, somehow, had to lie the stability and permanence demanded by scientific knowledge.

This situation gave birth to the notion of the universal. It required that a particular sensible thing, changeable by its very nature, should at the same time be a stable and abiding object of knowledge. It had to find in the sensible thing itself an object with necessity⁴ sufficient for scientific knowledge, that is, something that holds for all the particulars. The many and changing singulars had to be contained in a unitary whole that was unchanging, abiding, and all-embracing. Knowledge in terms of this unitary whole was knowledge *katholou*, knowledge bearing upon all the singulars as upon a distinctive whole. In that way the unitary whole was conceived as a "universal."

Since the universal, in the Aristotelian noetic, could not be an object outside the sensible thing,

¹ In the *Meno* (77A), Plato has *κατὰ δλον* to explicate adverbially his regularly used notion of *what something is*, in contrast to its singular instances. In the *Republic* (III, 392DE), *κατὰ δλον* expresses the consideration of a poem as a whole in contrast to its particular sections or verses. Against this background there is no difficulty in understanding the technical Aristotelian word *katholou* as basically a prepositional phrase with adverbial force. On this topic see Kurt von Fritz, *Philosophie und sprachlicher Ausdruck bei Demokrit, Plato und Aristoteles* (Leipzig, 1938), p. 65. The adverbial use, he notes, is by far the more frequent.

² Confrontations of these two sharply opposed trends in Greek thought may be seen in P. Merlan, "Isocrates, Aristotle and Alexander the Great," *Historia* (Wiesbaden), vol. 3 (1954), pp. 60-81; and in W. I. Matson, "Isocrates the Pragmatist," *The Review of Metaphysics*, vol. 10 (1957), pp. 423-427.

³ See Aristotle, *Anal. Post.*, II 19, 99b34-100b5; *De An.*, III 7-8, 431a14-432a9.

⁴ On the union of necessity and universality in the Aristotelian conception of scientific knowledge, see Suzanne Mansion, *Le Jugement d'Existence chez Aristote* (Paris, Louvain, 1946), pp. 18-107, esp. p. 95. On the difference between the Aristotelian setting for this question and the modern background that comes from Hume and Kant, see Francis H. Parker, "Traditional Reason and Modern Reason," *Philosophy Today*, vol. 7 (1963), pp. 235-244.

it had somehow to coincide with each of the singulars in which it was exemplified.⁵ Though a unitary whole, it had to be identical with each of the singulars in turn, and accordingly be predicable of each. This feature gave the universal its distinguishing characteristic within the general notion of "whole." The universal, consequently, was listed by Aristotle as one of the various meanings of that notion. It was "a whole in the sense that it contains many things by being predicated of each, and by all of them, e.g., man, horse, god, being severally one single thing, because all are living things."⁶

The universal, upon which scientific knowledge is based, is therefore for Aristotle something that is identical with each of the singulars in turn. That is why it can be predicated of each of them. That is why you may say, for instance, that Socrates is a man, and Callias is a man, and Coriscus is a man, and every other human individual is a man. Each of them is severally a "man," and "man," while a unitary whole, is each of them in turn. Each of the singular things instantiates one and the same universal. In each instantiation there is, each time, but one real object, able to be regarded in two different ways. In one way it is regarded as singular, in another way it is regarded as universal.

Because of this identity of the singular with the universal in each instance, both the formal and the material elements that compose the singular sensible thing are also found in the universal: "But man and horse and what are thus applied to singulars but universally, are . . . something composed of this particular formula and this particular matter, treated as universal."⁷ Like the sensible thing with which it is identical, the universal consists of the two contrasted constituents, matter and form. The only difference is that both thing and constituents are now regarded in a different way. In the one case they are regarded as singular, in the other case as universal.

This conception of the universal is spelled out

in further detail in the immediately following chapter of the Aristotelian treatise: "It is also clear that the soul is the primary substance and the body is matter, and man or animal is the compound of both taken universally."⁸ Here the soul, that is, the physical form,⁹ is mentioned as one constituent and is contrasted with the other constituent, the matter. The universal is characterized as the compound of both, regarded in universal fashion. The really existent man, "Socrates" or "Coriscus," is the compound of both constituents, and as such is a singular. The universal is also the compound of both, but of both as taken universally.

Universal and particular, accordingly, denote the same real thing,¹⁰ regarded in different ways. How is this possible? What grounds are there, either in the particular sensible things from which all human cognition originates for Aristotle, or in the human way of regarding them, that makes possible this conception of the universal? The difficulty is not brought out explicitly in the texts just considered. It is rather a problem that arises from them. But these texts do sketch in its general lines the Aristotelian notion of the universal. They allow one, therefore, to introduce a few preliminary points of precision that are necessary for clarity in a present-day discussion.

First, the Aristotelian universal is what is expressed by the predicate "man" in a sentence like "Socrates is a man." The same predicate "man" can likewise be united by the copula with any other individual such as "Callias" or "Coriscus." Callias is a man; Coriscus is a man. The predicate is one and the same in all the cases and is identical with each particular instance in turn. Each is the predicate, though the predicate itself as a universal remains a unit. The Aristotelian universal, accordingly, is a unitary whole that is expressed by a predicate like "man."

Second, the universal is clearly distinguished from the form that unites with the matter in both singular and universal. The form is only a part of

⁵ For the opposite view, namely, that the Aristotelian universal duplicates the sensible concrete reality, see Chung-Hwan Chen, "Universal Concrete, A Typical Aristotelian Duplication of Reality," *Phronesis*, vol. 9 (1964), pp. 48-57. This duplication, it is understood, "will cause the Aristotelian metaphysical system to break down" (*ibid.*, p. 57). The alleged dilemma that for Aristotle the real, because singular, is not knowable, and the knowable, because universal, is not real, has been revived recently by Whitney J. Oates, *Aristotle and the Problem of Value* (Princeton, 1963), pp. 180-183. Oates regards the solution given in *Metaph.*, M 10, 1087a15-21, as "Aristotle's extra, and unsuccessful, effort to solve it" (*ibid.*, p. 181). Aristotle's own statement of this aporia in regard to the first principles is found at *Metaph.*, B 6, 1003a5-15.

⁶ *Metaph.*, Δ 26, 1023b30-32; Oxford tr.

⁷ *Metaph.*, Z 10, 1035b27-30; Oxford tr., except "what" for "terms which" at b28.

⁸ *Metaph.*, Z 11, 1037a5-7; Oxford tr.

⁹ On the soul as form of the body, see *De An.*, II 1, 412a19-b6.

¹⁰ On the universal as given in sensation, see *Anal. Post.*, II 19, 100a17-b5; *Metaph.*, M 10, 1087a19-20.

the universal, just as it is only a part of the singular. The universal includes matter in its notion, while the form is the constituent that is distinguished from the matter. The form is a "this" (*tode ti*),¹¹ while the universal is regularly contrasted with a "this." The form is *ousia*, in fact primary *ousia*,¹² in the context of the *Metaphysics*, while the universal is not *ousia* from its viewpoint. Though "Form" and "Idea" coincide in Plato, in Aristotle form and universal are divergent. To conceive the Aristotelian universal as a Platonic Idea now functioning as an inherent form in matter would clash with the patent meaning of these Aristotelian texts.

Third, the Aristotelian universal, in its ordinary instances, is not what is represented by words like "humanity" or "whiteness."¹³ The universal is predicated of each individual in such a way that each one in turn is the universal. You say "Socrates is a man, Coriscus is a man, Callias is a man." But you do not say "Socrates is humanity, Coriscus is humanity, Callias is humanity." You may say "The wall is white, the snow is white, the metal is white," but not that each or any of them is "whiteness." "Humanity" or "human nature" is something composed of both matter and form, and is therefore obviously different from the form or soul of man. But how does it differ from the universal? Aristotle does not seem to have any interest in delving into this problem.¹⁴ Whatever his explanation might have been, however, notions like "whiteness" and "humanity" do not in their ordinary use correspond to his description of the universal.

The texts just considered, then, pinpoint the Aristotelian universal to what is expressed by a predicate like "man." They distinguish it definitely enough from what is expressed by notions like "humanity," and by the technical Aristotelian conception of "form" in contrast to matter. So

determined, the universal, although a unit, is identified in reality with each of many singulars in turn. For this reason it is predicable of each of them. The functioning of the universal, so understood, emerges readily enough from the structure of Aristotle's logic. The syllogism, that is, the process of reasoning, depends for its validity upon a middle term: "... if the universal goes, the middle term goes with it, and so demonstration becomes impossible."¹⁵ Because the middle term is universal with regard to the minor, and the major term is universal with regard to the middle, what is known independently of the major term becomes known thereby of the minor term. In this way human knowledge is increased through reasoning. Without the factor of universality, therefore, scientific knowledge would not be possible. As knowledge through conclusions inferred from premisses, science for Aristotle depends upon the universal.¹⁶

II

It is one thing, however, to show from the logical structure of reasoning that science depends upon the universal, and to require identity of universal and singular in order that scientific knowledge bear upon reality. It is quite another thing to show how this is possible. The sensible thing, from which all human knowledge takes its rise, is in every case singular. It is not a universal. It is something changeable. Yet it alone is what is directly attained in human cognition. How can knowledge of it provide knowledge of anything universal?

Obviously, where the knowledge is of the universal and the thing known is a singular, knowledge cannot be conceived as a mere replica of what is known, as a photographic reproducing or a cata-

¹¹ Cf.: "... that which, being a 'this', is also separable—and of this nature is the shape or form of each thing" (*Metaph.*, Δ 8, 1017b24–26; Oxford tr.); "... the formula or shape (that which being a 'this' can be separately formulated)" (*ibid.*, H 1, 1042a28–29); "... when ... the predicate is a form and a 'this', the ultimate subject is matter and material substance." (Θ 7, 1049a34–36); "... the nature, which is a 'this' or positive state towards which movement takes place" (Δ 3, 1070a11–12); "... form or essence, which is that precisely in virtue of which a thing is called 'a this'" (*De An.*, II 1, 412a8–9; Oxford tr.).

¹² "By form I mean the essence of each thing and its primary substance" (*Metaph.*, Z 7, 1032b1–2; Oxford tr.); "... there is a formula of it with reference to its primary substance—e.g., in the case of man the formula of soul" (*ibid.*, 11, 1037a28–29).

¹³ The two notions may be found confused in modern discussions, e.g., "Many philosophers have claimed that in addition to the objects met with in sense experience there exist entities of an entirely different and more esoteric kind, technically designated as 'universals'. According to this claim there are, in addition to such things as tables and white sheets of paper, the utterly different and less well-known objects tableness and whiteness." Morris Lazewitz, "The Existence of Universals," *Mind*, vol. 55 (1946), p. 1.

¹⁴ That a problem does arise from this distinction, however, may be seen from the careful and protracted consideration given it by St. Thomas Aquinas, *De Ente et Essentia*, c. II, ed. Roland Gosselin, pp. 10.20–23.7.

¹⁵ *Anal. Post.*, I 11, 77a7–8; Oxford tr.

¹⁶ Texts may be found listed in Bonitz, *Index Aristot.*, 243 Δ 279a22–25; Mansion, *op. cit.*, p. 94, n. 1.

loguing or a calculation of things. The mind is a life. It is a vital activity, and may be expected to elaborate its object, to work upon it, to represent it in a manner different from the here-and-now manner in which it exists in the real world. Can it do that without distorting or misrepresenting the thing known? Is there something about a sensible thing that enables it to be known in a way different from and wider than the way in which it actually exists in the real world?

These reflections suggest that two different sources have to be probed to determine the grounds of universality in Aristotle. One is the sensible thing that is known. What is the structure of the sensible object itself? Does its composition provide a basis for knowledge of it in universal fashion? Is there something about it, in its real existence, that allows it to be known as a universal? The other possible source is the mind's own activity. What type of cognitional activity can encounter a singular thing and know it as universal, without distorting or falsifying the object? Even though mind against an Anaxagorean background¹⁷ be granted a status above material mixture, is it thereby equipped to free a sensible object from the limitations of here and now? In saying that mind or soul is in some way all things,¹⁸ that the knower and the thing known are one in the actuality of cognition,¹⁹ does Aristotle mean that the knower confers his own way of being upon the thing known? But would that help in the present difficulty? The knower is an individual substance. He is not a universal. How could the knower's way of being help at all to explain universality?

These considerations, nevertheless, suggest an examination first of the sensible thing's structure, then of the activity of the human intellect, for the purpose of bringing to light any grounds for universality that may lie in each.

III

First, what is the structure of sensible things for Aristotle? The opening books of the *Physics* show

that every sensible substance is composed of two principles, matter and form. The matter, unknowable in itself, becomes knowable through analogy. As the bronze is to the statue and as the wood to the bed, so is the underlying nature in any sensible substance to its corresponding form.²⁰ The notion of an underlying nature within every sensible substance has resulted in the *Physics* from a long inquiry into the principles of change. The general conception of change, reached by analysis of observable instances like wood becoming a bed or bronze a statue, requires a subject to pass from one form into another. Change in substance, accordingly, means that an appropriate subject loses one form and acquires another. This subject is "the primary matter underlying each of the things that have in themselves the principle of motion or change."²¹ The matter, though in itself formless and in consequence entirely indeterminate and unknowable, has therefore become intelligible under the general notion of "subject," a notion acquired from observable substrates of change such as bronze and wood. But just in itself the matter contributes no intelligibility to the thing.

The other principle reached by the *Physics* is the form, in the sense of the first or basic formal principle. It is the fundamental knowable content of the thing, in contrast to the unintelligible matter and to the accidental characteristics. Through it are present the determinations that render a sensible thing knowable: "For only the form, or the object as having form, can be expressed in the concept; whereas the material element by itself cannot be expressed in the concept."²²

By actuating matter the form constitutes the singular thing. It can never be found in reality except as informing matter, that is, as in a singular thing and as physically repeated and physically distinct in each new instance of the singulars. But in itself is it singular? As a "this,"²³ it does stand in definite contrast to the universal. Yet in its own aspect as form it does not seem immediately to distinguish one sensible instance from another. Rather, the sameness of the form in all the many singulars is stressed:

¹⁷ See *De An.*, III 4, 429a18-22; b21-25. Cf. I 2, 405a13-17; b19-21.

¹⁸ *De An.*, III 4, 429b5-8; 30-31 (mind); 8, 431b20-29 (soul).

¹⁹ *De An.*, III 5, 430a19-20; 7, 431a1-2. Cf. I 4, 408b13-15.

²⁰ "The underlying nature is an object of scientific knowledge, by an analogy. For as the bronze is to the statue, the wood to the bed, or the matter and the formless before receiving form to any thing which has form, so is the underlying nature to substance, i.e., the 'this' or existent" (*Phys.*, I 7, 191a7-12; Oxford tr.).

²¹ *Phys.*, II 1, 193a29-30. It is called "ultimate subject" at *Metaph.*, Δ 8, 1017b24.

²² *Metaph.*, Z 10, 1035a7-9; tr. Richard Hope.

²³ See texts *supra*, n. 9.

... the begetter is adequate to the making of the product and to the causing of the form in the matter. And when we have the whole, such and such a form in this flesh and in these bones, this is Callias or Socrates; and they are different in virtue of their matter (for that is different), but the same in form; for their form is indivisible.²⁴

The use of the one Aristotelian term *eidos* for both "form"—which is a real physical principle—and for "species"—which is universal—may give rise to caution regarding the Oxford translation of *eidos* throughout this passage as "form." Yet in the passage *eidos* definitely means the form that is achieved in the matter through physical causality. The effect of the causality is undoubtedly a physical form. The concern is with a product that is made, that is begotten, in the real world. The passage does not bear upon a universal that is known in the mind. The *eidos* in question is a form "in this flesh and in these bones," and it constitutes singulars, namely Callias and Socrates. It is one constituent of the compound that is the real physical man.

The universal, on the other hand, is not one of the two physical constituents. It is the composite of both, taken universally. On account of one constituent, their matter, Callias and Socrates differ from each other. But in the *eidos* that has just been described, they are the same. Their form, under its own aspect of form and as expressly contrasted with matter, is not divisible. A sudden and unexplained change in the meaning of *eidos* from "form" to "species" in this passage could hardly be accepted, even though sameness in form results in sameness of species and is expressed in the Greek by exactly the same phrase.

The passage indicates that the form just as form does not immediately distinguish one singular from another. Human form (the soul), insofar as it is form alone, is the same in all instances of men. It serves to render all the instances human, to give them all the one specific nature. The distinction of the singular instances from one another follows upon the matter. Yet the form in itself is not a universal but a "this." To be in the real world the form of a sensible thing has to be singularized by matter. That is its essential requirement. When

real it has to constitute a particular man like Socrates or Callias. As the cause of being,²⁵ it imparts individuality to the matter and to the composite; for, unlike an absolutely separate form, it calls for realization solely in singular individuals. It requires that the thing of which it is the form be a singular.

In the real singular, accordingly, the form is the principle of determination and the principle that renders the thing knowable, though just of itself it does not immediately distinguish one singular from another. It is the immediate origin, rather, of the features common to them all. Has it not, therefore, the requisite condition for functioning as the ground of a knowledge of things that does not differentiate one singular from another, of knowledge applicable to them all? In its formal role it provides the source for common specific content and not immediately for discernment of singulars. This content extends equally and indifferently to all the singulars of the species. It makes knowable the composite, the entire thing. When a sensible thing is known in the way in which it is characterized by its form, it is indeed known as an object that includes matter. What is known is something composed of both principles, matter and form. Otherwise not the thing itself, but only its form, would be known. Yet the composite sensible thing is the origin of human cognition.

An object is characterized and made knowable by a form that just in itself does not distinguish its singulars. Has not the sensible thing, therefore, within its structure a principle that may readily serve as a ground for universal knowledge? Why may not the form characterize the matter in the known object in a way that leaves the object a composite with a knowable content common to all the singulars? This will require elaboration through the activity of the knower, for here the cognition does not reflect the way in which the matter singularizes the object in the real world. But it does retain the knowable content given the object by the form. As long as no new content is added by the mind's activity, the object will remain the same thing.²⁶ Only the universal way in which it

²⁴ *Metaph.*, Z 8, 1034a4-8; Oxford tr. Cf. 7, 1032a24-25. For the sense of same in species, see Bonitz, *Index Aristot.*, 218a43-52.

²⁵ *Metaph.*, Z 17, 1041b25-28; H 2, 1043a2-7.

²⁶ "For knowledge, like the verb 'to know', means two things, of which one is potential and one actual. The potency, being, as matter, universal and indefinite, deals with the universal and indefinite; but the actuality, being definite, deals with a definite object—being a 'this', it deals with a 'this'. But *per accidens* sight sees universal colour, because this individual colour which it sees is colour; and this individual *a* which the grammarian investigates in an *a*." *Metaph.*, M 10, 1087a15-21; Oxford tr.

is known will be new. The mind, in fact, has no natural tendency to attribute universality to the real things themselves, even when conscious that it is knowing them in universal fashion. In the object known there need be no falsification or distortion.

The universality, then, understood in this way, is not actually present in the real sensible thing. But it may be regarded as potentially present, because the intrinsic ground that allows it to be attained, the physical form, is actually present in the thing. Because of the form, the sensible thing can be known universally and can be defined.²⁷ The unitary character or sameness of the form as form in all the singulars will permit a common definition that gives expression to the total composite and will allow its content, matter as well as form, to be applied as a universal to all the singular instances.

This conception of the universal is not, of course, without difficulties. How, for instance, does it counter the thrust of Plato's *Parmenides* (131B): "If so, a Form which is one and the same will be at the same time, as a whole, in a number of things which are separate, and consequently will be separate from itself" (tr. Cornford)? Aristotle does not seem to feel the impact of this objection. The form is not a thing existent in itself. To ask if it is separate from itself is to place it first as existent in itself, instead of in matter. Perhaps the Stagirite could have answered that in the impossible supposition of a sensible form existing without matter it would allow only one individual in a species, since there would be nothing to multiply it.²⁸ Where being is explained ultimately by form,²⁹ as in Aristotle, the difficulty that each act of being means existential distinction is not felt. Within its own metaphysical framework, therefore, his notion of sameness in form does not seem to run into self-contradiction. But it can hardly hope to prove complete or satisfactory when questions in terms of existence are encountered.

²⁷ "Another question is naturally raised, viz., what sort of parts belong to the form and what sort not to the form, but to the concrete thing. Yet if this is not plain it is not possible to define any thing; for definition is of the universal and of the form." *Metaph.*, Z 11, 1036a26-29. Cf. "... for the formula that gives the differentiae seems to be an account of the form or actuality, while that which gives the components is rather an account of the matter." H 2, 1043a19-21; Oxford tr.

²⁸ So, for St. Thomas Aquinas (*Contra Gentiles*, II, 93) there can be only one separated substance in one species, because there is no matter to multiply singulars in it. For Duns Scotus, *Quaest. in Metaph.*, VII, 13, no. 21 (ed. Vivès, VII, 421b), the humanity that is in Socrates and the humanity that is in Plato would coalesce in reality if the individuating differences could *per impossibile* be struck away, because in that case there would be nothing to cause differentiation in their common nature "humanity." For Aristotle, "same in form" and "same in species" can readily coincide in the one Greek expression, since, on account of the type of causality exercised by the form, the singulars are one in species *because* they are one in form.

²⁹ This feature of Aristotle's thought was expressed neatly by Octave Hamelin: "La forme explique tout le reste et se suffit à elle-même." *Le Système d'Aristote* (Paris, 1920), p. 405.

³⁰ See *Metaph.*, Z 10, 1036a9-12; 11, 1036b35-1037a5; H 6, 1045a33-36; K 1, 1059b14-20.

³¹ II 19, 100a12-b5.

Another difficulty is that in the Aristotelian universal the material principle functions sufficiently to leave the object a composite of matter and form, but not sufficiently to render it singular. For Aristotle, however, matter may function in cognition differently from the way it functions in the sensible world. In the objects of mathematics it plays the role of intelligible matter, in contrast to sensible matter.³⁰ Within the Aristotelian framework, then, there need be no internal inconsistency in the stand that matter in a known object like the universal does not exercise all the functions it performs in the real sensible world. This consideration, however, throws the problem into the activity of the human intellect in its cognition of sensible things.

IV

A further ground for universality, then, has to be sought in the mind's activity. Through the well-known rout simile in the concluding chapter of the *Posterior Analytics*,³¹ Aristotle describes how cognition of the universal is attained. Attention is focused upon one among a number of indiscriminate particulars. For instance, the individual Callias is seen and understood as a man, as a unitary object under which all the other singulars may be brought, as "a single identity within them all." (100a7-8; Oxford tr.) The one explanation given is: "The soul is so constituted as to be capable of this process." (a13-14; Oxford tr.) The Stagirite speaks as though he is observing human cognition, noting the facts, and then saying that because as a fact the activity is taking place, the soul is accordingly capable of it.

Nor does the *De Anima* provide any further explanation of this capability on the part of the soul. In the soul there is a passive intellectual principle, and a corresponding active principle: "And in fact mind as we have described it is what

it is by virtue of becoming all things, while there is another which is what it is by virtue of making all things: this is a sort of positive state like light; for in a sense light makes potential colours into actual colours."³² The passage is extremely enigmatic. But on one point at least it is definite. In addition to becoming all things through cognition, the mind has a different phase of activity in the course of its grasping them. It is an activity illustrated by the simile of light making colors actually visible. The surfaces of sensible things, this would mean, are regarded as potentially colored. In the dark or the dusk, no differentiated colors appear. The light shines upon them, and the distinct colors become actually visible. Like light, this phase of mind is its unitary self (a22-23) and nothing more. But by its actual presence and activity it makes actual what was previously only potential in things.

The mind, then, is described as active insofar as in its knowing things it brings into actuality what was present in them only potentially. Does this consideration help in regard to universal knowledge? By reason of the form, what the sensible thing is offers a ground for universality. Can the activity of the intellect, bearing upon it like light, make it known in a way which, though actually individual,³³ is applicable to any other singular instance? The account is only by way of a simile. It does not seem to penetrate intrinsically into the workings of the intellect. Can it be regarded as anything more than the observation in the *Posterior Analytics* that such is the nature of the soul? The results of intellectual activity are carefully observed, and because they are found to be of this kind the intellect is declared to be of a nature able to effect them. The point is merely driven home with an illustration from the activity of light.

In particular, the Aristotelian account does not show how the mind is able to view an object in a way that allows matter to remain in the object's constitution without rendering it singular. The Stagirite stresses the fact that sensation, imagination, and intellectual cognition³⁴ receive the forms of sensible things without their matter. In this

rising scale of cognition there is obviously a question of degree.³⁵ In the real sensible world, matter limits the thing to a definite here and now. In external sense-perception it keeps the thing singular, but allows the one singular thing to be found in the cognitive activity of a number of different percipients. In the imagination it permits the things to be freed also from restrictions of time. In the specific universal it gives rein to knowledge above the limitations of space and time and above the conditions of contingency. In the more generic universals it opens out over a gradually increasing range. Aristotle, accordingly, sees differences in the way matter functions in the real world and in the various grades or types of cognition. In the cognitive activity of the knower the matter is observed to function in a manner appropriate to each grade. But this remains an observation. It is hardly a satisfactory explanation. Yet it is enough to show that for Aristotle the universal depends upon the activity of the human intellect as well as upon the form of the sensible thing.

V

The grounds for universality in Aristotle, therefore, are twofold. First, in every sensible thing there is a basic formal principle that, though individual, brings each instance into formal identity with all the other instances. Secondly, in human intellectual cognition there is an active principle that raises knowledge above the status of photographing or registering or cataloguing and actualizes what was only potential in the real thing. The unity indicated by the formal principle of sensible things is not something actual in the real world. In the real world each sensible thing is something apart from the others, and the form of each is physically distinct from the forms of the rest. But in knowing sensible things universally, the human intellect is able to grasp the concrete physical thing as characterized distinctively by its formal nature, to know it in a way that holds equally for all other instances of the form.

³² *De An.*, III 5, 430a14-17; Oxford tr.

³³ See text *supra*, n. 22. Cf. *Metaph.*, Z 10, 1036a6-8; 15, 1040a2-5; *Anal. Pr.*, II 21, 67a39-b3.

³⁴ For the three see *De An.*, II 12, 424a17-24; III 8, 432a9-10; III 4, 429b21-22 and 430a7-8 respectively.

³⁵ At *De An.*, II 12, 424a18-19 (Oxford tr.), a sense in general is described as "what has the power of receiving into itself the sensible forms of things without the matter." Yet at III 8, 432a9-10; Oxford tr., the phantasms are distinguished from the contents of sense-perception because—now in contrast to the latter—they are without matter: "for images are like sensuous contents except in that they contain no matter" (Oxford tr.). This would indicate in the phantasms a grade of immateriality not possessed by the contents of sense-perception. Further, mind is described (III 4, 430a7-8) as a power of attaining its objects without matter, implying a still higher grade of immateriality in cognition.

Aristotle has penetratingly observed the facts made manifest in the phenomena of scientific knowledge. He has analyzed them in his logic and thereby shown the functioning of the universal in human reasoning processes. The universal is there, as a datum. It is observable through reflection. To find in the Aristotelian treatises an adequate account of it in terms of its causes or grounds, however, becomes a disappointing endeavor. The physical form of the sensible object and the active phase of human intellection seem to be the grounds offered in the text. These are solidly based on careful observation. They avoid internal contradiction and offer a framework in which subsequent investigations have been carried on and in which future inquiries may be profitably pursued. But in their own location in Aristotle they can hardly be

expected to prove satisfactory. The protracted controversies on the universals throughout the middle ages, the continuing attacks from Nominalists in later centuries, the more recent approaches from the analysis of language instead of from the natures of things, as well as the widely divergent conceptions of soul and intellect in the long tradition of Aristotelian thought, all bear eloquent witness of the inconclusive status of the theme in the Stagirite's own text. Yet it would be rash to say that his account has lost its relevance. In point of fact, it still commands attention. Though incomplete, it remains a beacon light pointing out a way of inquiry, inviting further study, and providing a clearly defined framework in which fruitful discussions may continue to be carried on.³⁶ It is still a challenge.

Pontifical Institute of Medieval Studies, Toronto

³⁶ This paper was read at the meeting of the Western Division, American Philosophical Association, Chicago, May 1, 1965. Afterwards the chairman of the panel, Richard McKeon, called attention to the different ways in which the universal is regarded by Aristotle in his various *methodoi*. The consideration is important, and should always be kept in mind in the interpretation of the treatises, especially in regard to the ever variable way in which the universal functions in ethical matters. Nevertheless, a common feature throughout is that the universal is predicated of or belongs to the particulars. This can be verified for the theoretical treatises by a glance at Bonitz, *Index Aristot.*, 356b4-25, and for the *Ethics* by a passage like *E.N.*, V 7, 1134b18-1135a8. The present paper has been concerned with a metaphysical inquiry into the grounds in general that make possible the common predication throughout the areas covered by the various *methodoi*.

AMERICAN PHILOSOPHICAL QUARTERLY

Edited by
NICHOLAS RESCHER

With the advice and assistance of the Board of Editorial Consultants:

William Alston
Alan R. Anderson
Kurt Baier
Lewis W. Beck
Roderick M. Chisholm
L. Jonathan Cohen
James Collins
James M. Edie
Peter Thomas Geach

Adolf Grünbaum
Carl G. Hempel
John Hospers
Raymond Klibansky
Ernan McMullin, S. J.
Benson Mates
John A. Passmore
Günther Patzig
Richard H. Popkin

Wesley C. Salmon
George A. Schrader
Wilfrid Sellars
Alexander Sesonske
J. J. C. Smart
Manley H. Thompson, Jr.
James F. Thomson
G. H. von Wright
John W. Yolton

VOLUME 3/NUMBER 3

JULY 1966

CONTENTS

- | | | | |
|--|-----|--|-----|
| I. M. J. SCOTT-TAGGART: <i>Recent Work on the Philosophy of Kant</i> | 171 | IV. R. L. PURTILL: <i>Moore's Modal Argument</i> | 236 |
| II. KURT BAIER: <i>Moral Obligation</i> | 210 | V. RODERICK M. CHISHOLM AND ERNEST SOBA: <i>On the Logic of "Intrinsically Better"</i> | 244 |
| III. JAEGWON KIM: <i>On the Psycho-Physical Identity Theory</i> | 227 | VI. DWIGHT VAN DE VATE, JR.: <i>Other Minds and the Uses of Language</i> | 250 |
-

AMERICAN PHILOSOPHICAL QUARTERLY

POLICY

The *American Philosophical Quarterly* welcomes articles by philosophers of any country on any aspect of philosophy, substantive or historical. However, only self-sufficient articles will be published, and not news items, book reviews, critical notices, or "discussion notes."

MANUSCRIPTS

Contributions may be as short as 2,000 words or as long as 25,000. All manuscripts should be typewritten with wide margins, and at least double spacing between the lines. Footnotes should be used sparingly and should be numbered consecutively. They should also be typed with wide margins and double spacing. The original copy, not a carbon, should be submitted; authors should always retain at least one copy of their articles.

COMMUNICATIONS

Articles for publication, and all other editorial communications and enquiries, should be addressed to: The Editor, *American Philosophical Quarterly*, Department of Philosophy, University of Pittsburgh, Pittsburgh, Pennsylvania 15213.

OFFPRINTS

Authors will receive gratis two copies of the issue containing their contribution. Offprints can be purchased at low cost through arrangements made when checking proof.

SUBSCRIPTIONS

The price *per annum* is six dollars for individual subscribers and ten dollars for institutions. Checks and money orders should be made payable to the *American Philosophical Quarterly*. Back issues are sold at the rate of two dollars to individuals, and three dollars to institutions. All correspondence regarding subscriptions and back orders should be addressed directly to the publisher (Basil Blackwell, Broad Street, Oxford, England).



I. RECENT WORK ON THE PHILOSOPHY OF KANT

M. J. SCOTT-TAGGART

§1. INTRODUCTION

I have been able to limit the scope of this review through the existence of two excellent surveys of recent work on Kant that have been made by de Vleeschauwer.¹ Presupposing an acquaintance with these I have been able to deal only in a passing way with books that appeared before 1960, and so to deal more fully with shorter articles that go back as far as the *Festschriften* of 1954. My treatment is necessarily selective and is largely confined to those areas where there is some unity of discussion.

I shall preface a few remarks on general trends in Kantian studies. Although for several decades there was a fair degree of commerce between the Continental and the English speaking schools, the last twenty or thirty years has seen a considerable divergence of both interests and methods, which is partly unnecessary, and totally undesirable. About the English speaking scene I shall say little at this stage. Here, despite two recent studies of the first *Critique*,² the attention of most significant Kant scholars has been focused more upon Kant's ethical views. Largely due to the influence of Paton, we have had not only two studies of the *Foundations*, but also commentaries on both the *Critique of Practical Reason* and the *Metaphysic of Morals*,³ as well as a host of free or detailed treatments of various Kantian topics.⁴

On the Continent things are managed rather differently. Interest here is still centered on the first *Critique*. While English commentary stems, through

Paton and Kemp-Smith, from the work of Adickes and Vaihinger, things happened in Germany in and around 1924 that have scarcely begun to influence the English scene. Most significant of these were the works of W. Wundt,⁵ Nicolai Hartmann,⁶ and Heimsoeth,⁷ which provided a point of departure for a school which has come to be known as the ontological or historical-ontological school. While not all authors can be reckoned as belonging to this school, most can be shown to be influenced by it, or, at the least, to share features in common with it, and it is to these common features that I wish here to draw attention.

The first feature comes out well in de Vleeschauwer's revision of some of his earlier views.⁸ While still insisting that we must look at Kant through the question "Is metaphysics possible?", he now points out that the historian must look at Kant in terms of both individuality and continuity. Previously we have looked to Kant for a revolutionary originality, but we must now look at him as the product of his time. We must notice, for example, how the *objectum materiale* of the *Critique*—i.e., metaphysics—corresponds exactly to the Wolffian scheme, with the Analytic corresponding to *metaphysica generalis*, and the Dialectic to *metaphysica specialis*, and, noticing this, we shall be led to renew our study of the *Critique* "not as a divided totality, but as a unity that comprehends within itself an entire tradition of thought and systematization."⁹ This suggestion, that we must confront Kant with the rationalist metaphysicians of his time—most

¹ *A Survey of Kantian Philosophy* (1957); *Etudes kantienues contemporaines* (1963). Another review is to be expected shortly in *The Monist*. Full details of all recent work cited can be collected from the bibliography.

² Bird, *Kant's Theory of Knowledge* (1962); Wolff, *Kant's Theory of Mental Activity* (1963). Cf. §6.

³ Ross, *Kant's Ethical Theory* (1954); Duncan, *Practical Reason and Morality* (1957); Beck, *Commentary on Kant's Critique of Practical Reason* (1960); Gregor, *Laus of Freedom* (1963).

⁴ Silber, Singer, and Schrader deserve special mention here.

⁵ *Kant als Metaphysiker* (Stuttgart, Enke, 1924).

⁶ "Diesseits von Idealismus und Realismus," *Kant-Studien*, vol. 29 (1924).

⁷ Cf. §§2-3.

⁸ "Wie ich jetzt die Kritik der reinen Vernunft Entwicklungsgeschichtlich lese" (1963). A comment on his earlier *La Deduction transcendente dans L'Oeuvre de Kant* (1934-7) and *L'Evolution de la Pensée Kantienne* (1939).

⁹ *Ibid.*, p. 356.

notably Baumgarten, Wolff, and the early Kant—has been acted upon in the articles of Heimsoeth and Tonelli, as well as in the books by Martin and Delekat. It describes an approach to Kant that has produced three decades of historical studies of Kant's work that are informed by scholarship and illuminated by imagination.

There are signs that this approach is overreaching itself, even though it does provide a necessary corrective to the Neo-Kantian bias.¹⁰ One of its most important results is to show that the contrast of "Precritical" and "Critical" views is an exegetical fiction—a polemical device for eliminating unwanted parts of the *Critique*. The continuity in Kant's thought is so great that it leads Heimsoeth to use, and not only to mention, the word "Critical" within quotation marks. But while these studies show that an understanding of much of the *Critique* is contingent upon an understanding of Kant's earlier work (as ethical studies are showing how understanding of the *Foundations* is contingent upon an understanding of the later work), there is a danger of minimizing those features of Kant's thought that made the Neo-Kantian movement possible. These features may not have sprung into existence between 1770, or 1768, and 1781, but they came into existence during Kant's lifetime, and their isolation in a revisionary interpretation is a legitimate task. They must even find their place in an historical account of Kant's work, and we must not allow the historical approach of the ontological school to guarantee the historical validity of their results.

A determining ground of this possible exaggeration of the metaphysical motives in Kant's development lies in the conscious reaction of recent commentary against Neo-Kantianism, and this on philosophical as well as upon historical grounds. This phenomenon appears in one of the manifestos that Martin has published for the German school:

Neo-Kantianism restricted the systematic task of philosophy to theory of knowledge, and consequently extended this restriction to the philosophy of Kant and to its interpretation. The ontological school contests this restriction. It contests in general that philosophy exhausts itself in theory of knowledge, and it

contests in particular that the philosophy of Kant exhausts itself in this way.¹¹

This claim is problematic. I shall not say anything about the general claim that philosophy is not reducible to epistemology, although it means that a lot of German commentary is being written by philosophers with an axe of their own to grind. This is also true in a straightforward sense of English commentary, but here the convictions to be saved are Neo-Kantian convictions. So far as Kantian exegesis is concerned, the approach to Kant is original in that it ceases to be obsessed (as the Neo-Kantians were) with the relations of Kant to Newton and Hume,¹² while the leading interest of English authors is still centered on these relations. The approach declares Kant's interest in founding a "practical-dogmatic metaphysics," and not merely in laying a safe foundation for geometry and Newtonian physics. One English commentator who shares this view is Silber. He argues for the primacy of Kant's ethical views,¹³ and maintains:

That Kant should direct metaphysics to the objects of freedom, God, and immortality, only to carry out the inquiry into these objects in terms of the ideas of the soul, the world, and God, clearly shows that Kant presents the inquiry into the moral ideas of reason as a genuinely metaphysical and speculative inquiry as well as a moral investigation.¹⁴

As a feature of the work of the ontological school, this is not by any means as widely shared as the first, yet it combines with it to produce an interest in Kant that is not restricted to the Deduction and the Analogies. As a result parts of Kant's work that were previously of only incidental interest have moved closer to the middle of the stage, and these are studied in the context of all the material, unpublished in Kant's lifetime, that is now so readily available. As a further result, the question of Kant's view of things in themselves has again become the turning point against which different attitudes can be gauged.

§2. THINGS IN THEMSELVES

I have said that the ontological school is distinguished from the Neo-Kantian movement by

¹⁰ Cf. Lehmann, "Kritizismus und kritische Motive in der Entwicklung der kantischen Philosophie" (1956-7).

¹¹ "Die deutsche ontologische Kantinterpretation" (1958) in *Gesammelte Abhandlungen und Vorträge* (1961), p. 105.

¹² Newton, but not Hume, is returning to favor in the work of Delekat, Schmucker, Lehmann, and Hildebrandt.

¹³ "The Context of Kant's Ethical Thought" (1959). The argument from the extended history of Kant's intention to write a *Metaphysics of Morals* is valuably supplemented by Beck, cf. *Commentary* (1960), pp. 5-18.

¹⁴ "The Metaphysical Importance of the Highest Good as the Canon of Pure Reason in Kant's Philosophy" (1959), p. 243.

mutually complementary historical and philosophical interests. Commenting on previous commentary, Martin says that:

the question was constantly posed of whether Kant was able to speak of things in themselves. Being answered negatively, the things in themselves were either excised altogether from the Critical philosophy, or else reinterpreted so that nothing but the word remained.¹⁵

The ontological school is distinguished from this style of commentary by an absolute insistence on Kant's commitment to things in themselves. The later treatments are mostly developments of positions taken by Heimsoeth in his influential thesis on the *Metaphysische Motive in der Ausbildung des kritischen Idealismus* (1924), which is in many respects to be taken together with the equally important article on *Persönlichkeitsbewusstsein und Ding an sich in der kantischen Philosophie* (1924).¹⁶ In the first of these articles Heimsoeth exhibits the conscious reaction to the Neo-Kantian doctrines that I have mentioned, claiming to show that:

the Critical limitation of knowledge (in particular the withdrawal of spatio-temporal experience from actual being in itself as a mere knowledge of appearance) is determined by certain basic metaphysical convictions.¹⁷

He first of all argues that intellectual intuition is not an artificially constructed concept that emerges at the same time as the distinction between sensibility and understanding, but "it signifies for Kant the primary representation of knowledge: as a pure and spontaneous mental intuition."¹⁸ But a knowing subject intuits only what he himself "makes." Heimsoeth shows himself aware of the difficulties of the metaphor of making by using quotation marks, but whatever its difficulties it plays an undeniable part in Kant's thinking on intellectual intuition, which is often called "productive intuition," and

which can be traced back to the writings of the middle fifties. Man, however, as a finite intelligence has spontaneity limited by receptivity: and all knowledge which one finite substance possesses of another is the product of both mental spontaneity and of a real affection through something beyond itself.¹⁹ If this is correct, then it would follow that Kant always maintained a contrast of things in themselves and things as they appear, and this is what Heimsoeth claims. It existed before 1770 as the contrast between internal and external properties. With the *Dissertation* the external properties acquire their own principle of unity and form the phenomenal world: but the things in themselves remain. Three characteristic quotations may illuminate this bare account:

(1) finite beings cannot know other things out of themselves, for they are not their creators, so that it is only the mere appearances that they are able to know *a priori*,²⁰ (2) man is the *principium originarium* of the appearances,²¹ and (3) as the noumenon within ourselves is related to the appearances, so is the highest intelligence related to the *mundi intelligibilis*.²²

The first two quotations date from later than 1781, and show the connection between a priority and spontaneity, while the last illustrates the contrast of Divine and human understanding, and dates from the silent decade. Occurring with these dates they leave untouched the problems of showing the role of spontaneity in the Precritical period, and of the role of receptivity in the Critical period, but they nevertheless give content to a vigorous and extremely suggestive article that breathes new life into the old problem of Kant's view of things in themselves.

Some of the distinctive insights are developed by Herring.²³ After giving an historical account of the different views on the role of things in themselves, Herring starts from the distinction between finite and infinite intelligences to which Heimsoeth draws attention, and goes on to develop a theory

¹⁵ "Die Probleme einer ontologischen Kantinterpretation" (1952), in *Gesammelte Abhandlungen und Vorträge* (1961), p. 82.

¹⁶ These and other writings from 1924-6 have been collected in *Studien zur Philosophie Immanuel Kants* (1956), which also contains one previously unpublished essay, while a second collection of articles dating from between 1917 and 1960 is published as *Studien zur Philosophiegeschichte* (1961).

¹⁷ *Op. cit.*, p. 191.

¹⁸ *Ibid.*, p. 192.

¹⁹ Heimsoeth speaks of knowledge as a *Seinsrelation*, and links it with causality. There would appear to be more than a Lockean basis for this, but what more is involved I am not sure.

²⁰ Refl. 6048, Ak. XVIII. 433.

²¹ Refl. 6057, Ak. XVIII. 440.

²² Refl. 5109, Ak. XVIII. 91.

²³ *Das Problem der Affektion bei Kant* (1953).

that involves both a double affection theory, and also a double aspect account known to us through Paton.²⁴ Thus, for example, it is said that the distinction of things in themselves from appearances concerns:

neither one and the same object of experience to which two contradictory predicates are applied, nor does it concern two ontologically distinct objects; it is, rather, one and the same object that is regarded under two different aspects, as object of experience and as object of pure thought.²⁵

There is a great deal to agree with in Herring's article, but I should like to single out one aspect for special consideration, and for comparison with the views of Bird. Both men maintain that there is a distinction to be drawn between the transcendental object and things in themselves.²⁶ Bird does so on the grounds that, by means of the transcendental object, Kant poses the problem that:

the thought of an object in general is presupposed in all our empirical knowledge, and it is this thought that stimulates Kant's problem about the meaning of our notion of an object. This notion is precisely not the thought of any intelligible object, but only the idea of certain objective features of our knowledge and experience.²⁷

In maintaining such a view Bird fails to give an account of those passages where Kant uses the term "transcendental object" and "thing in itself" as synonymous.²⁸ These passages have given rise to an orthodox view which is reported by Wolff:

using Kant's terminology more precisely than he was accustomed to do, we may say that he began with the concept of a transcendent object = *x*, and then shifted to the concept of a transcendental object = *x*. The former is merely the concept of the thing in itself, but the latter is the concept of the ground of the

unity of a manifold of representation in one consciousness.²⁹

It is to be noticed, however, that Wolff is not altogether consistent about this, since strong support is also given to the alternative view that the terms are always used synonymously by Kant.³⁰ The "orthodox" view to which I refer is that which accepts that there is equivocation between "transcendent object" and "transcendental object," and treats this as linked with a distinction between "Precritical" and "Critical" views: perhaps also in the strong sense of the patch-work theory.³¹ It is the "Critical" view that Bird is maintaining, revealing his to be essentially a revisionary interpretation of Kant. Herring, however, accepts that there is an equivocation in the term "transcendental object," but thinks of this equivocation as essential to it. He concludes that the affecting object is the transcendental object, which has the two transcendental aspects of being either appearance or thing in itself.³²

I find that I cannot fully agree with any of these opinions. I might mention the reason for this, since it does illustrate the usefulness of one of the features of modern European Kant exegesis. The question has never been asked: What is the origin of the concept of the transcendental object? If we ask this question, we find that the concept is used in the *Critique* in one of the ways that Kant used the concept of God in his earlier writings. Like Leibniz, Kant found difficulty in the idea of external relations, and it is this that is troubling him when, asking the question "How is it possible that there should be outer intuition in a thinking subject?" he answers that no man can answer, because in order to answer it we should have to possess knowledge of the transcendental object.³³ This passage quite clearly needs to be interpreted in the light of the passage in the *Dissertation* where he says that:

²⁴ *Kant's Metaphysics of Experience* (1936), vol. II, op. 422.

²⁵ Herring, *op. cit.*, p. 82.

²⁶ Both Bird and Herring use A253 to sustain their case, but where Herring supports this with A358, Bird uses A493-6=B521-4, which is used by Herring rather differently (cf. p. 93).

²⁷ *Kant's Theory of Knowledge* (1962), p. 80.

²⁸ Adickes, in *Kant's Opus Postumum* (1920), gives the following list: A358, 361, 366, 372, 379-80, 393, A46=B63, A180-1=B236, A277=B333, A288=B344, A478=B506, A538-40=B566-8, A565=B593, A698=B726. Compare Herring, p. 81 where the import of the list is obscured.

²⁹ *Kant's Theory of Mental Activity* (1963), p. 314.

³⁰ *Ibid.*, p. 136.

³¹ Classic expositions are those of Adickes, Vaihinger, and Paton. The distinction is not, in fact, paired with the patch-work theory, because Kemp-Smith adopted the view that the terms were used synonymously, in which opinion he is joined by Aebi and Cassirer.

³² Herring, *op. cit.*, pp. 83-84.

³³ A392-3.

the human mind . . . senses things external to it only through the presence of the one upholding common cause; and space, which is the universal and necessary condition, sensitively apprehended, of the co-presence of things, can therefore be entitled [the] *omnipraesentia phaenomenon* [of the general cause].³⁴

Already we find Critical diffidence, for Kant speaks of this as "overstepping the limits of apodictic certainty befitting metaphysics," and as "pushing out into the open sea of mystical enquiries." From this passage, however, the issue can be traced farther back through the *Habilitationsschrift*³⁵ to Kant's first published work,³⁶ where it is held that only through God does our knowledge acquire relation to an object. In the *Critique* the ontological guarantee gives way to an epistemological problem; but the nature of this problem, and the call for the transcendental object, can be defined through consideration of the Precritical role played by God. This is a topic which has been collecting attention in its own right. Schmucker has provided us with an excellent account of Kant's different positions in relation to the three main theological arguments at four stages of Kant's development: in 1755, 1762, 1765, and 1770.³⁷ Kopper has argued for the importance of theological considerations in Kant's development through the writings from the *Beweisgrund* to the *Opus Postumum*,³⁸ although his interpretation of the *Opus Postumum* has been challenged by Schulze.³⁹ Valuable studies have also been contributed by Redman⁴⁰ and Henrich,⁴¹ while Klausen⁴² and Rust⁴³ take us into the later writings. The validity of Kant's criticisms, particularly of the ontological argument, which has summoned so much recent English controversy, is here subordinated to the question of the implications of those criticisms, and these implications are shown to be of the utmost importance for an understanding of Kant's development.

A second major problem over things in them-

selves has always been whether we are entitled even to speak of things in themselves, i.e., whether denying that they can be known, we are asserting something meaningful, let alone true, if we say (or imply) that they exist. Traditional discussions show that advance is only likely to be made through an understanding of Kant's use of the modal concepts.⁴⁴ Schneeberger's excellent monograph on these concepts is therefore to be welcomed.⁴⁵ Of singular importance is the concept of possibility which, he reminds us, is built up out of the concept of agreement; and, that we may have different species of possibility depending upon the quantity of statements with which we require agreement. In the case of empirical possibility we may demand consistency with all the laws of an existing body of science, or with all the laws of an ideal science, or with either of these in conjunction with a set of factual conditions. In the case of real possibility we require consistency with the necessary conditions of experience, i.e., require that it be a possible experience. These are both questions of "external possibility." In the case of logical, or "internal possibility," we require no more than self-consistency, and, because this concept is less specific than that of real possibility, it licenses talk of objects which are not objects of a possible experience.

More work will have to be done before this line of approach can afford us an altogether acceptable introduction to things in themselves, but, quite apart from this Kantian context, Kant's views on possibility deserve development. The Humean "test" for possibility in terms of imaginability is horribly circular, and our constant use of it blurs the Kantian distinction between internal and external consistency.⁴⁶ It is because of my objections to the Humean account that I find Schneeberger's account more helpful than a distinction between "imaginable" and "conceivable" with which

³⁴ Ak.II. 409-410.

³⁵ Ak.I. 413-415.

³⁶ Ak.I. 20-21.

³⁷ "Die Gottesbeweise beim vorkritischen Kant" (1969).

³⁸ "Kants Gotteslehre" (1955-6).

³⁹ "Zu Kants Gotteslehre" (1956-7).

⁴⁰ *Gott und Welt, Die Schöpfungstheologie der vorkritischen Periode Kants* (1962).

⁴¹ *Der ontologische Gottesbeweis* (1960).

⁴² *Das Problem der Erkennbarkeit der Existenz Gottes bei Kant* (1959).

⁴³ "Kritisches zu Kants Religionskritik" (1955).

⁴⁴ Cf. Herring, *op. cit.*, p. 70: "What Kant denies is not the reality of the transcendental object as such, but only its empirical reality."

⁴⁵ *Kants Konzeption der Modalbegriffe* (1952).

⁴⁶ There is more than a punning connection between this use of "internal" and "external" and their use in relation to internal and external properties, but this has not yet been investigated.

Stenius moves into the same area.⁴⁷ Stenius is concerned with the move from "The categories are *a priori* valid of appearances" to "The categories are *a priori* valid no farther than appearances": the first of which summarizes Kant's critique of empiricism, while the second is the foundation of his critique of rationalism. Stenius holds that Kant makes the move in the *Critique* through his view that everything conceivable must also be imaginable; and, although I find this contrast unrevealing, I would take it to be related to Kant's move from the insistence of the *Dissertation* that the categories must not be subject to sensible conditions to the position of the *Critique* where he insists that the *a priori* validity of the categories can only be established in relation to such sensible conditions. The "real use" of the understanding gives way to the Critical understanding of "real possibility," and the fallacy of subreption gives way to the doctrine of schematism.⁴⁸ It is through a philosophical understanding of these historical moves that I would expect progress to come in our understanding of the Kantian position over things in themselves.

§3. THE EARLY WRITINGS

Erdmann at one time suggested that a decision over Kant's philosophical motivation in 1769 carried with it a decision over the foremost positions of the *Critique*,⁴⁹ and the suggestion is well founded. The year 1769 is a crucial one in the Kantian development, and it is certainly, and perhaps trivially true that philosophical and historical differences of opinion are always linked with one another. It is therefore significant that such writers as Heimsoeth, Martin, Delekat, and Tonelli call upon 17th and 18th century scenes that are much more richly populated than those familiar to less recent commentators. The point is worth emphasizing because there is a formal similarity between the position of the ontological school and the position adopted by Erdmann concerning the influences determining Kant in 1769: both emphasize the Antinomies.

Martin has shown the peculiar importance of the first two Antinomies for the ontological school,⁵⁰

and, more recently, Heimsoeth has been devoting considerable attention to them. In the first of a group of three articles he cited the normal passages to show the influence of the Antinomies upon Kant's development,⁵¹ but he argues beyond them, and against the position of Paulsen, that "the idea of an empirical-sceptical period determined by Hume and others in the development of Kant's thought has become untenable."⁵² He argues for this by taking the framework so often used by Neo-Kantian authors, but traces Kant's idea of a progression from dogmatism through scepticism to criticism not to the middle term of Humean scepticism, but instead to the scepticism of Bayle in his famous article on Zeno. The merest glance at the *Reflexionen* is sufficient to show the way in which Kant arranged his thought within the framework of positions labeled by the names of Plato and Leibniz, Aristotle and Epicureus, and it is by developing such hints that Heimsoeth is able not only to show the defects of a naive Neo-Kantian view, but also to throw new light on the Antinomies by locating Kant in the main streams of Western Philosophy. One of the particularly important results of his enquiry is the essential connectedness of the Second Antinomy and the Second Paralogism that it reveals. The following passage from Samuel Clarke gives an idea of the way in which material simplicity and psychological unity were connected in the eighteenth century mind:

... suppose three, or three hundred, particles of matter, at a mile or any given distance from one another, is it possible, that all those separate parts should in that state be one individual conscious being? Suppose, then, all these particles brought together into one system, so as to touch one another, will they thereby or by any motion or composition whatsoever, become any whit less truly distinct beings than they were at the greatest distance? How, then, can their being disposed in any possible system make them one individual conscious being?

The purely material question Heimsoeth deals with mainly in connection with the *Monadologia Physica* and the question of the relation of soul to matter is treated through the *Träume*. Granted the

⁴⁷ "On Kant's Distinction between Phenomena and Noumena" (1963), pp. 236-240.

⁴⁸ But not altogether. Reason comes in as a third faculty to hold the field.

⁴⁹ *Reflexionen Kants zur Kritik der reinen Vernunft*, vol. II, (1884), sect. XXIV.

⁵⁰ *Kant's Metaphysics and Theory of Science* (1955), pp. 42ff; "Zu den Voraussetzungen und Konsequenzen der Kantischen Antinomielehre" (1940), reprinted in *Gesammelte Abhandlungen und Vorträge* (1961).

⁵¹ Ak. XII. 254; Ak. XVIII. 69; Ak. IV. 338, etc.

⁵² "Vernunftantinomie und transzendente Dialektik in der geschichtlichen Situation des kantischen Lebenswerkes" (1959-60), also in *Atom, Seele, Monade* (1960), p. 8.

importance of the Antinomies, and given the amount of work that has now been done on their Kantian history,⁵³ this relating of the two questions offers hope of important advance because, if we cease looking for what is changing rather than for what is constant, the question of how the soul is present in the world is revealed as one of the questions most constantly in Kant's mind throughout his life.

Another neglected Precritical work has always been the *Allgemeine Naturgeschichte*, and this has also been given an Heimsoeth treatment in his shorter article on the First Antinomy.⁵⁴ Kant's view of a definite origin combined with "Entwicklung in dem Abflusse der Ewigkeit" is related once more to traditional views and to its immediate ancestors in Baumgarten and Knutzen. This time I was less satisfied, perhaps because I have always wanted this view related to the writings of the same period where Kant, for example, says that although every contingent event has its determining reason, not every determining reason has, of necessity, its consequence.⁵⁵ Paneth has provided a second, and very different sort of essay devoted to Kant's cosmological views.⁵⁶ He shows the relationship between Kant's views on the milky way and those of Thomas Wright, delineates the way in which Kant eliminated the Newtonian God that held the fixed stars apart, and argues that despite the frequent linking of their names, there is a great deal of difference between the positions of Kant and Laplace.

Heimsoeth's example has been of considerable importance for shaping the structure of contemporary Kant commentary, and one of the most important results has been that we are now more inclined to view the transition from early to late periods in terms of continuity rather than discontinuity.⁵⁷ We may well, however, disagree with Heimsoeth's own description of this result when he says that:

the conceptual world of the practical-dogmatic metaphysics, as projected and prepared by Kant in his later writings, stands in unbroken continuity with the conceptual world of the precritical period.⁵⁸

While not contesting the continuity, and also not contesting the importance of the metaphysical elements in the earlier part of the story, we may well believe that Heimsoeth's interpretation of the later writings leads him to exaggerate the metaphysical elements at the expense of the Critical elements in Kant's development. This thesis has been argued by Lehmann.⁵⁹ He does not want to reintroduce Hume into the picture, but he wishes to insist that at an academic level Kant's thought should be represented as moving between the two poles of Newtonian physics and the metaphysics of the schools. Beneath this academic polarity, however, Lehmann sees Kant's development in terms of a factor that has not been previously emphasized, namely Kant's changing views on anthropology. Kant is described as moving from a "cosmological naturalistic" picture of man in the *Allgemeine Naturgeschichte* through 1772/3 to the "spiritual" picture of man in the *Träume*, and finally to the "critical" picture of man as legislative for the phenomenal world and practically engaged within it. This vein of Kant's thought is a neglected one, but the terms have a stubborn indefinability that makes it doubtful whether the approach could also be a useful one.⁶⁰

Another force has been introduced into this field by Tonelli. Because of my stumbling Italian I find his two major works frustrating, but they have been reviewed by de Vleeschauwer.⁶¹ There is a shorter article which advertises the results of the forthcoming second part of the second of Tonelli's books in which Tonelli argues that the decisive move made by Kant in 1769 was the separation of sensibility and understanding, and the explanation of space and time as pure forms of intuition is not the driving force leading to this separation, but merely

⁵³ Apart from Heimsoeth, see especially Carl Siegel, "Kants Antinomielehre im Lichte der Inaugural-Dissertation," *Kant-Studien*, vol. 30 (1925-6), pp. 67-86, and H. Feist, *Der Antinomiedanke bei Kant und seine Entwicklung in den vorkritischen Periode* (Borna-Leipzig, 1932).

⁵⁴ "Zeitliche Weltunendlichkeit und das Problem des Anfangs" (1960).

⁵⁵ Ak. I. 396, 408-409.

⁵⁶ "Die Erkenntnis des Weltbaus durch Thomas Wright und Immanuel Kant" (1955-6).

⁵⁷ Compare the reviews by Heidemann (1956), whose own *Spontaneität und Zeitlichkeit* (1958) is deeply influenced by Heimsoeth, and by Knittermeyer (1957-8) and Wagner (1961-2).

⁵⁸ "Chr. Wolffs Ontologie und die Prinzipienforschung I. Kants" (1956), p. 64.

⁵⁹ "Kritizismus und kritische Motive . . ." (1956-7).

⁶⁰ Cf. Weiler, "Kant's 'Indeterminate Concept' and the Concept of Man" (1962).

⁶¹ *Dall' estetica metafisica all' estetica psicoempirica* (1955), and *Elementi metafisici e metodologici in Kant* (1754-1768), vol. I (1959). See de Vleeschauwer's "A Survey of Kantian Philosophy" (1957), pp. 140-142, and "Etudes kantienues contemporaines" (1962-3), pp. 78-81.

a consequence of it.⁶² Such a theory, while extremely attractive, is also extremely difficult to carry through, because the distinction of sensibility and understanding is the meeting point of a cluster of other distinctions that it manages, often illegitimately, to hold together. Kant, for example, is inclined to argue:

space and time are intuitions, and are therefore to be ascribed to sensibility, and being ascribed to sensibility are therefore to be accounted representations of how things appear rather than how they are.

In the face of such complexity attention is likely to devolve onto one of the subordinate distinctions, but Tonelli's work promises not only to face but also to master the complexity. Two things seem to me to be particularly significant here. The first lies in the importance of Baumgarten's aesthetic theory, of which Tonelli takes note, and of which no one is better qualified to speak. The second is the way in which Tonelli's intimate acquaintance with the eighteenth century scene enables him to reintroduce Hume into the picture: not directly as some Neo-Kantians would have it, in determining the views of the *Negative Grössen*, but indirectly, through Basedow's *Philalethie*, as significant for the position of the *Träume*. This sort of influence is so much more plausible than insistence upon direct influence, the question of which has been recently reargued by Wolff.⁶³ In a slight amplification of Kemp-Smith's views, Wolff argues that although Kant knew the *Essays* through Sulzer's translation of 1752-4, he was only acquainted with the *Treatise* through the quite extensive quotation from it contained in Beattie's *Essay on the Nature and Immutability of Truth*. It would be wrong to lay too much stress on this, since Humean views were well known to Kant, even if the words of the *Treatise* were not.

Tonelli's scholarship has overflowed into a series of articles that are, one feels, to be regarded as by-products of his main activities. They are historical rather than philosophical analyses, which provide the material for synthesis with other Kantian elements. Such is his treatment of the *Prize Essay*, where Kant's separation of the methods

of metaphysics and geometry is placed against a rich background of seventeenth and eighteenth century authors, with their opinions, and with Crusius and Tetens singled out as particularly relevant to Kant's own position.⁶⁴ In another article Kant's criticism of the idea of substance is placed against the background of similar criticism launched by other authors,⁶⁵ and in a third the same task is done for the concept of infinity.⁶⁶ The same style is used in two other articles with greater success. These are devoted to examining the previous occurrences of some of the new terms that appeared with a systematic role in Kant's thought after 1769. Some of the earlier ones can be traced to Leibniz and Locke (*perceptio purus*, *intuitio*), others to Lambert (transcendental, transcendent), or Darjes (analytic, dialectic).⁶⁷ Tonelli points out the way in which so many of these terms rarely occur in the eighteenth, but belong to the German Aristotelian tradition of the seventeenth century; and hence through them Kant could emphasize his difference from the Wolffian school while not introducing a technical vocabulary that was totally alien to his readers. This sort of work, while not overtly dramatic, is extremely useful. It can have its dramatic moments, however, and Kahl-Furthmann's investigation of the reversal of meaning that has occurred in the words "subject" and "object" since the middle ages is a fascinating example of this. His argument carries well to the conclusion that:

in the transference of the term originally relating to the knowledge side to the side of the object, the insight into the contribution of the subject in the constitution of the object is implicitly expressed.⁶⁸

and, it is interesting that this index of the change from pre-Kantian to post-Kantian pictures of the world should be traceable to the early seventeenth century.

I have been relating this discussion to 1769 because today, even more than at earlier times, the *Dissertation* is seen as a Critical work, and the most decisive break in Kant's thought as that between 1768 and 1770. But this is in part illusory, since the break between 1766 and 1768, although not as

⁶² "Die Umwälzung von 1769 bei Kant" (1962-3), p. 369.

⁶³ "Kant's Debt to Hume via Beattie" (1960).

⁶⁴ "Der Streit über die mathematische Methode in der Philosophie der ersten Hälfte des XVIII Jahrhunderts, etc." (1955).

⁶⁵ "Critiques of the Notion of Substance Prior to Kant" (1961).

⁶⁶ "La question des bornes de l'entendement humain etc." (1959).

⁶⁷ "Das Wiederaufleben der deutsch-aristotelischen Terminologie etc." (1963); "Der historische Ursprung der kantischen Termini 'Analytik' und 'Dialektik'" (1962). Cf. works on the categories cited below.

⁶⁸ "Subjekt und Objekt etc." (1953), p. 334.

extensive as that after 1768, laid the foundation for the conceptual apparatus that was developed in the *Dissertation*. It was in 1768 that Kant dropped his view that spatiality was a phenomenon dependent upon the *commercium substantiarum*, and so made possible its disengagement from the noumenal world. His argument for this is that from incongruent counterparts.

There have been several studies devoted to this phenomenon. The first, by Pears,⁶⁹ has been developed by Mayo.⁷⁰ These studies, although not explicitly concerned with the way in which Kant made use of incongruent counterparts to establish a theory of space, deal with many of the peculiarities of the phenomenon through outlining the number of contingent statements which lead us to say that we are reversed right to left when we look in a mirror. Mayo's article is mainly valuable for orientating ourselves within the Kantian problem, and seems to me to contain only one mistake, which may well be a simple slip. This is that, on Mayo's account, any three dimensional object will be incongruent with its mirror image,⁷¹ whereas for Kant this is only true of three dimensional objects which exhibit no plane of symmetry. Two objects are incongruent if they will not fit into the same spatial boundary. This is important in relation to a second article that deals specifically with the Kantian use of the phenomenon. Peter Remnant argues that the existence, or possibility, of incongruent counterparts does not entail an absolute space.⁷² His argument is not without difficulties. He supposes Kant's argument to be that if a single hand were created, then, if a handless human body were next created, the hand would either be a right or a left hand, and therefore would previously have been a right or a left hand. The "previously" he believes to be a mistake. He implicitly points out that a handless body is virtually left-right symmetrical, and therefore he supposes the symmetry entirely disturbed by the right sides of our handless bodies being red, and the left sides being green. If, now, "right" meant "on the red side of the body," then God could make a single hand either right or left by next creating a red-on-the-introspectively-right-side body or creating the mirror image of this, i.e., a red-on-the-introspectively-left-side body.

Instead of the body being colored, we could just as easily suppose a handless body created that had been conditioned to report our right by "left" and our left by "right." Thus Remnant concludes that "in a universe which contains nothing but a single hand, it would not just be empirically undecidable whether that hand were right or left; it would be strictly indeterminate."⁷³

I am not sure that this supposed indeterminacy refutes Kant's claim that the existence of incongruous counterparts is impossible without the existence of an absolute space: I should think it could be construed as essential to Kant's argument rather than as a refutation of it. Perhaps Remnant plays into Kant's hands. For let us suppose that God begins his creation with a single hand: Remnant would surely not deny that he has performed a different act of creation from that which he would have performed if he had created its counterpart. But according to Remnant's argument this cannot mean "He created a right instead of a left hand," since this remains indeterminate until God creates other things. But we can surely say categorically that whichever hand God created, it occupied a different region of space from that which its counterpart would have occupied: this must follow from the definition of incongruence, according to which incongruent figures are those which cannot occupy the same space. Therefore if God cannot choose between the two hands by asking "Which way round shall I place it?" but only by asking "Where shall I put it?" it seems that if the question is to be significant then space must exist antecedently to objects.

I do not think that if the question is to be significant it must entail the existence of space, nor, indeed, do I think that the question is significant. But if (1) Remnant allows that one hand will occupy a different space from that which its counterpart would have occupied—and how can this be denied?—(2) he also allows, as he does with some diffidence,⁷⁴ that there is not any "intrinsic difference" between the two hands, and (3) he does not allow God to choose by means of the question "Which way round?" then I do not see that he can avoid Kant's conclusion that God will have to ask "Where?" and thus, apparently, pre-

⁶⁹ "The Incongruity of Counterparts" (1952).

⁷⁰ "Incongruity of Counterparts" (1958).

⁷¹ *Ibid.*, pp. 109–110.

⁷² Peter Remnant, "Incongruent Counterparts and Absolute Space," *Mind*, vol. 72 (1963), pp. 393–399.

⁷³ *Ibid.*, p. 399.

⁷⁴ *Ibid.*, pp. 394, 399.

suppose an absolute space. The solution to the problem lies in denying (2); i.e., maintaining that there is an intrinsic difference between a right and a left hand, and because of this God can choose by asking "Which?" Space must prevent me from dealing with this here, as also with the uses that Kant makes of incongruent counterparts in the *Dissertation* and the *Prolegomena*, although a recent article by Lange,⁷⁵ replying on Kant's behalf to an older article by Reidemeister, throws a certain amount of light on this topic that has remained dormant since Vaihinger's synoptic treatment.

§4. THE AESTHETIC (I)

In this section I wish mainly to consider just two authors, whose work, because not laced with scholarship, is liable to be overlooked by the Kant scholar, and yet which was clearly done with Kant in mind. These are the investigations of Strawson⁷⁶ and Quinton,⁷⁷ which contain, to my mind, the most significant contributions to our understanding of Kant's theory, or theories, of space and time. Their importance lies mainly in connection with Kant's description of space and time as "pure intuitions": they do not aid our understanding of the phrase "form of sensibility," and thus not our understanding of the connection of these two descriptions.

Although Kant moves easily between the substantial terms "Vorstellung" and "Anschauung" and their verbal forms "vorstellen" and "anschauen," it is clear that the transition is not an easy one. But if we take the terms substantively the main outlines are clear: intuition and concept are two species of representation, and Kant makes it clear that an intuition is to be distinguished as a singular representation.⁷⁸ He is, however, often unclear whether his contrast between singular and general is to be construed as the contrast between singular and plural, or as the contrast between particular and universal. Which of these contrasts we think more fundamental will determine, for example, the way in which we interpret the first two space arguments, where Kant argues that space is "essentially single."⁷⁹ Without begging

this question, it is, however, undeniable that "There can be no more than one space" and "There can be no more than one time" are integral parts of Kant's position. They are integral parts in this way: they are necessary conditions for describing space and time as "forms of the sensible world." By this it is meant that space and time will, if these statements are true, provide a necessary system of relations in which it is possible that every particular should be located, and, through this location, obtain a relation to every other particular. In order to say that space and time are forms of the sensible world, however, Kant needs to conclude not only that it is possible for every particular to be located in space and time, but also that this is necessary. This necessity is in part expressed and argued for in the first two space arguments, where Kant argues that space is a "condition of the possibility of appearances." These connections are exhibited in the following passage:

Space, therefore, is an absolutely first formal principle of the sensible world, not only for the reason that the objects composing the universe cannot be phenomena save through the concept of space, but especially for the reason that by its essence space has to be single, embracing absolutely all outer sensibles, and so it constitutes a principle of totality, i.e. of a world which cannot be part of another.⁸⁰

This is as much as I may do here toward indicating the importance to the Kantian theory of the statements (1) "There is one space," (2) "There is one time," and (3) "Every object has a spatio-temporal location." They are conjunctively necessary and sufficient for the statement (4) "There is one world." It is (1)-(3), however, that are investigated by Strawson and Quinton.

It is the third of these that is of particular concern to Strawson. *Individuals* is a book that has grown out of Strawson's earlier objections to Russell's theory of descriptions, investigating the presuppositions of our using, and of our being able to use, referring expressions. The second part of the book, dealing with the different ways in which subject and predicate terms are introduced into language, is intimately related to the first, although mainly as presupposing it. The first part can there-

⁷⁵ "Über den Unterschied der Gegenden im Raume" (1958-9).

⁷⁶ *Individuals* (1959).

⁷⁷ "Spaces and Times" (1962).

⁷⁸ Ak. IX. 91.

⁷⁹ A25=B39.

⁸⁰ Ak. II. 405.

fore be treated, although it ought not to be treated, independently of the second. This first part consists of four chapters: on Bodies, Sounds, Persons, and Monads.

The second of these, although of great importance in connection with the Analogies, is relevant to the problems of the Aesthetic in two main ways. It is particularly relevant to the Kantian thesis that "is outside us" and "is in space" are, in one sense of "outside us," equivalent expressions. Although "All objects outside us are in space" may very well be an analytic statement, in one modern use of this expression, there is also good reason for describing it as synthetic and *a priori*.⁸¹ Second, the chapter is mainly concerned with re-identification, and by showing that we could not have a use for "A is qualitatively unlike B, although numerically identical with it" unless we also had a use for "A is qualitatively like C and is numerically identical with it," and that we could not have a use for this last statement unless we also had a use for "A is qualitatively like D, but is numerically different from it." Strawson shows how many of the problems involved with re-identification are solved by means of answering questions about individuation. But since he also shows that the spatio-temporal conditions for identifying particulars are only realized through our ability to re-identify particulars, the circle involved can be used to show the mutual dependence of the Aesthetic and Analytic.

The argument of the chapters on Bodies and Monads is concerned with showing the way in which the systems of spatial and temporal relations provide the necessary means for making identifying references to individuals: that these cannot be mediated purely descriptively in the manner of Leibniz. The philosophical issues here have also been clarified by the able discussions of several other philosophers.⁸² The importance of spatiality and temporality for making identifying references, insisted upon in the rejection of the identity of indiscernibles, can be said to be something that Kant recognized, for he says in 1770 that:

all our intuition is bound to a certain formal principle under which alone anything can be apprehended by the mind immediately, that is, as singular, and not as conceived discursively through general concepts.⁸³

If, then, experience is to be experience of particulars, as it must be if we are to have synthetic judgments, we have good reason for saying that Kant was saying that space and time are necessary conditions of experience in that they are necessary for individuating particulars.

I am not sure that the significance of Kant's rejection of the identity of indiscernibles has yet been appreciated. I might hint at this importance by remarking that in its context the passage just quoted is an argument from the falsity of the identity of indiscernibles to the impossibility of intellectual intuition. The problem about the identity of indiscernibles is essentially a problem about external relations, and so to concentrate on it is to concentrate upon the distinction between things in themselves (the sum of the internal properties) and phenomena (which exhaust themselves in external relations), and through this upon Kant's distinction between transcendental realism and transcendental idealism. Some of these connections are dealt with by Herring in an article that may well be of use in showing how Strawson's arguments may be integrated, to the benefit of our understanding, into Kantian exegesis. I cannot agree, however, with his conclusion that Kant does not contest the identity of indiscernibles as such, "but solely its illegitimate use to argue by analogy from the subjective representation of an object to its constitution in itself."⁸⁴ Kant's argument against the identity of indiscernibles in the Amphiboly⁸⁵ is a dressed up version of an argument that long antedates *die kritische Wendung*, and the terminology of sensibility and understanding.⁸⁶ This earlier argument provides the basis for the Kantian discontent, and it cannot be described in the terms that Herring employs.

I have suggested that Kant maintains that every particular must have a spatio-temporal location because without this it would not be a particular.

⁸¹ Cf. §7.

⁸² Cf. papers under the title "Identity of Indiscernibles" by A. J. Ayer, *Philosophical Essays* (London, 1954); M. Black, *Mind*, vol. 61 (1952), pp. 153-164; D. Pears, *Mind*, vol. 64 (1955), pp. 522-527, and N. Rescher, *The Journal of Philosophy*, vol. 52 (1955), pp. 152-155, as also G. Bergmann, "The Identity of Indiscernibles and the Formalist Definition of 'Identity'," *Mind*, vol. 62 (1953), pp. 75-79, and N. L. Wilson, "The Identity of Indiscernibles and the Symmetrical Universe," *Mind*, vol. 62 (1953), pp. 506-511.

⁸³ Ak. II. 396.

⁸⁴ "Leibniz' principium identitatis indiscernibilium und die Leibniz-Kritik Kants" (1957-8), p. 399.

⁸⁵ A263-4=B309-20.

⁸⁶ Ak. I. 409-410.

But this is not Kant's only reason for saying that every sensible particular must be in space and time. For Kant an element *A* formed part of the world made up of further elements *B . . . N* if and only if *A* were related in some way to every other element of *B . . . N*. The systems of spatial and temporal relations, as systems in which every point of each system is uniquely related to every other point of that system, provide a means whereby every element, through occupying a position in those systems, can be related to every other element. The systems of spatial and temporal relations are means of conferring unity upon the world.⁸⁷ This, I believe, is Kant's main interest in the singleness of space and time, and he is only incidentally interested in the fact that through them we can uniquely identify any particular element: although for this latter feature we must also, it seems, presuppose the statement that every point of space and time is uniquely related to every other point of space and time.

The relationship of being spatially connected is ordinarily a transitive one, so that if *A* is connected with *B*, and *B* with *C*, then *A* will also be connected with *C*. But it seems that every object that I perceive will be spatially related to my body, and therefore this, through the terms of the Copernican revolution, will act as a conduit for spatially connecting all the particulars that I perceive. The most contingent feature of this situation is my association with one particular body, and it is through an implicit challenge to this feature that Quinton has recently thrown light upon the necessity of there being only one space. We might suppose, to start with, that a person is in perceptual relation with the world through two bodies, and so is perceiving the world from two different positions. It is only, of course, if the two bodies stand in spatial relation to one another that they will act as a conduit for connecting all perceived particulars, and Quinton considers a situation where this connection is broken. He supposes there to be a person who falls asleep in England, dreams of waking in the midst of a lakeside community where he spends a normal day before falling asleep there and, at that moment, wakes in England. He argues that if there were regular alteration, with waking in one world always coordinated with falling asleep in the other, we should have no reason to speak of the lakeside world as a fictitious world; and he takes the story far enough, to my mind, to establish its

freedom from inconsistency. Under these conditions we should speak of "two spaces," and the normal transitivity of spatial connectedness would have broken down.

Quinton's argument seems to me to demonstrate the falsity of the Kantian claim that "we can represent to ourselves only one space."⁸⁸ I do not think it affects the validity of Strawson's argument. In relation to this it would only emphasize the importance of perceptual contexts in relation to individuation. But while it demonstrates the falsity of a Kantian claim, this only serves to highlight the more correct and important aspects of his other claims. In his early work Kant was inclined to rule out the possibility of a plurality of worlds by means of two premisses (1) there is only one God, (2) God always works through the Leibnizian principle of providing a maximum coherence for his handiwork.⁸⁹ If there were more than one God, then Kant concedes not only the possibility but also the necessity of there being more than one world.⁹⁰ In the *Critique*, however, man adopts the position in relation to world unity that had been the Precritical prerogative of God, and it is therefore important to note that it is only by disturbing our concept of a unitary person, as a complex of body and mind, that we have been able to create the possibility of there being more than one world; and this is therefore analogous to the Precritical consideration of the possibility of there being more than one God. To follow through the implications of this argument (which I cannot do here) is therefore to locate oneself in the essential Kantian framework of one space, one world, and one consciousness.

I have considered only one space, and not one time. Quinton considers it impossible that there should be more than one time, and although he is indubitably correct in holding that there must, in any story that we tell, be a single subjective time order, it is not the case that our concept of an objective time order would always remain totally undisturbed. This is important in relation to the Analogies. Even in the story which Quinton tells, the temporal correlations of happenings in the two worlds would be a relatively *ad hoc* affair, employing incidental rather than essential features of our normal methods of establishing whether *A* happened before, after, or at the same time as *B*. If we extend Kant's epistemological basis, and consider not one man but a group of men, then it seems quite possible to make a distinction between "sub-

⁸⁷ Cf. Ak. II. 398.

⁸⁸ A25=B39.

⁸⁹ Ak. I. 25.

⁹⁰ Ak. I. 414-415.

jective" and "objective" time that will enable us quite easily to speak of "two objective times." I have myself, in an article on private languages appearing shortly, commented briefly and inadequately on the extra assumptions involved in relating objectivity to society rather than to an individual, and a story of the required kind for speaking of two times has recently been told by Swinburne.⁹¹

§5. THE AESTHETIC (II)

Some of the other work that has been done on the Aesthetic I comprehend only imperfectly, and my only safety against misrepresentation lies in brevity. There is, first of all, the work of Kaulbach, who has published a series of articles dealing with the concept of space. The main work traces Kant's distinction of two different kinds of space out of the early writings into the *Critique*, where the concept of motion is developed as a concept that enables us to unite Kant's transcendental idealism with his empirical realism.⁹² There are two shorter essays that provide a useful introduction: one that deals with a distinction that may be drawn between two ways of regarding space,⁹³ and another that identifies motion as the essence of the concept of synthesis.⁹⁴ There is a fourth work which has only just come to hand, which traces Kant's concept of motion into the *Opus Postumum*, and which promises, for me at least, to shed retrospective light on the other articles.⁹⁵

What makes assessment so difficult is that there is still no general agreement on Kant's relation to Newton and Leibniz. Delekat gives a view that is subscribed to by many. He maintains that Kant's *Gemüt* is formally equivalent to Newton's *sensorium Dei*, and argues for this by placing an 18th century environment around the dispute. He suggests various analogies designed to make the translatability of these terms into one another more appealing. One of these is Shaftesbury's use of "*sensus communis*," which helps very little indeed, since Shaftesbury took his term from Juvenal, and uses it to point a contrast between "the sense of the

common people" and "the sense for the common people," which has very little relevance to the perceptual context. More appealing is Augustine's use of the same phrase, which is made relevant by Delekat's interpretation of "*sensorium Dei*," of which he says that:

as there is a central part of the human brain in which, when we perceive, the sensations from the particular senses come together to produce a picture of the whole of what is perceived, and so provide the possibility of our being present to things independently of the place and time in which we find ourselves: so it is with God.⁹⁶

That is, just as I am present to, although not in, this typewriter when I perceive it, so God is present to space and time although not in them. God's *sensus communis* and my own resemble each other in everything except that his is not supplied with messages from various afferent nerves; i.e., God's intuition is *originarius* and not *derivatus*.

Having given this view of a *sensus communis*, Delekat is in a position (a) to formally identify Kant's *Gemüt* with a human, and Newton's *sensorium Dei* with a divine *sensus communis*, (b) to argue that the content of Kant's space is the same as the content of the Newtonian space, and (c) to say, therefore that "Kant changed nothing in the material determinations of the concepts of space and time, but he located them in the human mind rather than in the *sensorium Dei*."⁹⁷

Delekat's interpretation of Newton is somewhat idiosyncratic. Martin, who starts from the orthodox view of Newton as that given by Clarke, according to which space is an attribute of God, comes to a very different conclusion. Contesting (a) he also contests (b), saying that "for Kant [like Leibniz] space is merely a complex of relations."⁹⁸ In this conflict of views it is difficult to isolate the separate issues involved, even in so far as this merely involves the locating of the disagreement as one about the interpretation of Kant, or of Newton, or of Leibniz. The position is made even more complicated when notice is taken of the fact that the Aesthetic and the Analytic present different views. Schrader has called attention to this, saying that:

⁹¹ "Times" (1965).

⁹² *Die Metaphysik des Raumes bei Leibniz und Kant* (1960).

⁹³ "Geist und Raum" (1959).

⁹⁴ "Das Prinzip der Bewegung in der Philosophie Kants" (1963).

⁹⁵ *Der Philosophische Begriff der Bewegung* (1965).

⁹⁶ *Immanuel Kant* (1963), p. 33.

⁹⁷ *Ibid.*, p. 62, cf. Lehmann, "Kritizismus und kritische Motive . . ." (1956-7), p. 39.

⁹⁸ *Kant's Metaphysics and Theory of Science* (1955), p. 37.

intuition presents space in its absolute and non-relational character, whereas understanding is concerned with the relational aspect of space and time.⁹⁹

The mutual dependence of Aesthetic and Analytic is therefore something over which we need to become very much more clear before any final resolution to the problem is possible. Here, apart from the work of Strawson and Kaulbach that I have mentioned, the present article by Schrader provides some novel views, for he suggests that in the Aesthetic all that Kant does is "give several examples of sense perception in which the awareness of space and time is involved," whereupon "he generalizes and arrives at the conclusion that all perceptions are in space and/or time."¹⁰⁰ This inductive argument is combined with psychological analysis to show various features of the "presentationally immediate," and then, in the Analytic, Kant "attempts to show that what intuition reveals about space and time is presupposed in empirical judgments and, further, that it must be presupposed. This constitutes the deduction offered for space and time."¹⁰¹ I find I cannot agree with this position, but the article is highly recommendable as one of the clearest to be found on Kant's spatial views.

A related topic can also be introduced through Delekat's position. For one of the questions which Delekat puts before Kant is this:

Are space and time forms of intuiting, or forms which can themselves be intuited? If we understand the concept of intuition in the sense of *intuitus originarius* . . . then space and time are forms of what is intuited. But how is it if we regard them as pure intuitions of mankind? In this case are they *only* forms of intuiting, or are they *also* forms of what is intuited?¹⁰²

The question inevitably reminds one of the famous description of Kant's argument in the *Aesthetic*, namely:

Space and time are *a priori*, because necessary and universal,¹⁰³ and if *a priori* then subjective, and *only* subjective.

It was this description that led to the most acrimonious of all Kant debates: the controversy between Adolf Trendelenburg and Kuno Fischer at the end of the last century. Trendelenburg had held that:

even if we concede the argument that space and time are demonstrated to be subjective conditions which, in us, precede perception and experience, there is still no word of proof to show that they cannot at the same time be objective forms.¹⁰⁴

History has so far given the verdict to Trendelenburg. Despite Fischer's advocates including such men as Arnoldt and Caird, the final statements on the controversy have been those of Vaihinger and Kemp-Smith, so that today Trendelenburg's position is almost that of an unquestionable truth.¹⁰⁵ The position that he wished to maintain was that space and time could be transcendently real as well as transcendently ideal, and although this was disputed, common ground was found in the consistency of asserting that space was both empirically real and transcendently ideal. Delekat is therefore going in the face of all received opinion when he says that Kant's attempts to reconcile "form of intuiting" with "form of what is intuited" simply cannot be carried through.¹⁰⁶

I believe that the whole of this controversy could well come up for reappraisal, and with Delekat taking sides against both parties it is perhaps time that it did. There is further reason for this reappraisal. In apparent ignorance of the earlier controversy, Lotz has recently been arguing an extremely strong version of the Trendelenburg thesis. He believes that disputes have been based upon the unquestioned (!) assumption that the absolute reality of space is inconsistent with its transcendental ideality, and so he proposes to question this assumption.¹⁰⁷ So far this would appear to be the same ground as the earlier dispute, but Lotz goes on to argue not only that the absolute reality of space is quite consistent with its transcendental ideality, but also that it is a presupposition of it. He proposes to give:

⁹⁹ "The Transcendental Ideality and Empirical Reality of Kant's Space and Time" (1951), p. 530.

¹⁰⁰ *Ibid.*, p. 521.

¹⁰¹ *Ibid.*, p. 529.

¹⁰² *Op. cit.*, p. 62.

¹⁰³ Notice that even this part of the formula is contested by Schrader.

¹⁰⁴ *Logische Untersuchungen* (1862), p. 163.

¹⁰⁵ Cf. Körner, *Kant* (1955), pp. 37-38 and Schrader, *op. cit.*, p. 516: "it seems that a realistic explanation of the *a priori* is not only compatible with Kant's transcendental method of proof, but actually is more consistent with it."

¹⁰⁶ *Op. cit.*, p. 64.

¹⁰⁷ "Die Raum-Zeit Problematik in Auseinandersetzung mit Kants transzendentaler Ästhetik" (1954), p. 31.

a kind of transcendental deduction of the absolute reality of space from its transcendental ideality.¹⁰⁸

His arguments for this are as simple as the uncovering of what I take to be its errors would be complicated and instructive. My main objection here must be that the earlier part of the thesis, i.e., the view that space and time might possibly (rather than necessarily) be both transcendently ideal and real is not argued for. Without such argument, and the clarification of the terms that it would bring, it is impossible to reach an opinion on the later part of the thesis. And such an argument would be the reopening of an old case.

§6 THE ANALYTIC

Two recent English studies of Kant's account of objectivity will provide the core of this section, but before I come to them I should like to mention some shorter articles that deal with different aspects of the Analytic. The chapter on schematism has collected some attention,¹⁰⁹ but a great deal more has been paid to the metaphysical deduction. One line of investigation here concerns the relation of the tables of judgments and of the categories. The view that Paton held in *Kant's Metaphysics of Experience* on the core of the deduction, where Kant holds that "the same function which gives unity to the various representations in a judgment, also gives unity to the mere synthesis of various representations in an intuition,"¹¹⁰ has been under attack by Smart¹¹¹ and defended by Paton,¹¹² while Grayeff has proposed an intermediate position¹¹³ and Vuillemin has independently proposed a version that resembles the thesis maintained by Smart.¹¹⁴ The question at issue is that of the relation of formal and transcendental logic, and not that of the success of the deduction in providing us with a complete and adequate list of categories; and, it may, as Paton insists, be put in terms of the problem of the relation of the distinction between

analytic and synthetic *unities* to the distinction between analytic and synthetic *judgments*. This is one of the most fundamental questions that we may ask about the *Critique*, and it is important to realize that any answer to it will be consistent with the view, maintained by Delekat, that the actual list of categories "is not immediately derived from logic, but is developed from a consideration of the Leibniz-Wolff ontology."¹¹⁵ This view, that the table of judgments is engineered to produce the required list of categories, is not really contestable, but new light has been thrown upon the view by the researches of Heimsoeth in two smaller articles subordinated to the investigations of *Chr. Wolffs Ontologie und die Prinzipienforschung I. Kants* (1956).¹¹⁶ The article does not attempt to give a full account of all the *termini ontologici* of the Wolffian metaphysics, but, starting from the principles of contradiction and of sufficient reason, Kant's move toward his own late standpoint is traced through the *Negative Größen* to the "Tafel der Einteilung des Begriffs vom Nichts" and the highest principle of all synthetic judgments. The essay then organizes itself around the Kantian categories, for which the Wolffian analogues are discovered and discussed, and a strong case is made for regarding the categories as retaining some part of their original ontological role within the phenomenal realm, rather than in their supposed relationship to Newtonian physics.

We may turn from these articles to some that deal with the refutation of idealism. The most surprising of them is an excellent article by Turbayne, who upsets all our previous views on the relationship of Kant and Berkeley.¹¹⁷ It was, of course, allowed that Berkeleian and Cartesian idealism were simply positions created by Kant as contrasts to his own, but Turbayne argues to my complete persuasion that (1) Kant could have been acquainted with Berkeley's *Three Dialogues* and *De Motu*, (2) there is a point by point parallelism between Kant's and Berkeley's accounts of the

¹⁰⁸ *Ibid.*, p. 32.

¹⁰⁹ Rotenstreich, "Kant's Schematism in Its Context" (1956); Walsh, "Schematism" (1957-8); Schaper, "Kant's Schematism Reconsidered" (1964).

¹¹⁰ A79=B104.

¹¹¹ "Two Views on Kant and Formal Logic" (1955).

¹¹² "Formal and Transcendental Logic" (1957-8).

¹¹³ "The Relation of Transcendental and Formal Logic" (1959-60).

¹¹⁴ "Reflexionen über Kants Logik" (1960-1).

¹¹⁵ *Immanuel Kant* (1963), p. 75.

¹¹⁶ "Zur Geschichte der Kategorienlehre" (1952), and "Zur Herkunft und Entwicklung von Kants Kategorientafel" (1962-3). Cf. also Tonelli's "L'origine della tavola dei giudizi e del problema della deduzione delle categorie in Kant" (1956), and "La tradizione della categorie aristoteliche nella filosofia moderna sino a Kant" (1958).

¹¹⁷ "Kant's Refutation of Dogmatic Idealism" (1955).

external world, (3) Kant knew that there was, but (4) distinguishes his position from that of Berkeley by his theory of the *a priori* nature of space. Turbayne's argument is a model of clarity throughout, and my only objection to it is a slight tendency to underestimate the force of (4); but, a more accurate estimation of its force, while slightly disturbing the parallelism, would endorse the general conclusion that Kant was fully aware of his general agreement with the Berkeleian position.

If this is true, then even more attention than heretofore must be paid to the polemical nature of the Refutation of Idealism that Kant inserted into the second edition of the *Critique*. This is a point which, on other grounds, is insisted on by Lehmann in an informative article that relates Kant's Refutation to the *Lose Blätter* of the eighties and nineties, and to the *Opus Postumum*.¹¹⁸ The old question of the relation of first and second editions has become, with the greater availability of Kant's later work, part of the very much larger question over the development of Kant's thought after the publication of the three Critiques. Lehmann is prominent among those who evidence a growing concern with the *Opus Postumum*,¹¹⁹ but several other authors are also concerned with the work and with Adickes' interpretation of it;¹²⁰ and we may expect, what is desirable, that the next decade will see a host of studies devoted to tracing particular concepts from the three Critiques into the *Opus Postumum*, and that these studies will have important results for our understanding of the earlier works. A small indication of the utility of such studies is provided in the present context by an article of Müller-Lauter,¹²¹ in which he follows up the expanded base for the Refutation of Idealism that is provided by the *Nachlass*, and to which Lehmann draws attention. Incidentally concerned to show the relation of the first and second edition refutations, he argues against Vaihinger that the two do not conflict, but

the main interest of the article lies in the exposition of the second edition refutation. It is shown that when Kant speaks, in the second sentence of the Refutation, of all perception presupposing something permanent "in der Wahrnehmung,"¹²² the reference to perception involves reference to passivity, and passivity to externality, and externality to spatiality.¹²³ These are connections which Kant himself makes in the *Nachlass*,¹²⁴ and I would take them to show the metaphysical affiliations of the Refutation. They are not so taken by Müller-Lauter.¹²⁵ They do, at least, introduce new factors into the discussion.

A book that moves in this same area of Kant's views on objectivity, and which, if retrograde from the point of view of recent German commentary, is often exciting and persuasive, has been provided by Bird. Positively, and it is only with this aspect that I shall be concerned, the book is a study of Kant's word "appearance" and of its relations to the categories and to the concept of a thing in itself. It deals first of all with the problem of how we might reconcile Kant's conflicting claims about appearances, for these are said to be "objects, spatial, and distinct from our ideas, and yet they are also representations, mere modifications of the mind, and in us."¹²⁶ The problem is dealt with through a distinction between transcendental and empirical beliefs. The crucial point that Bird wishes to make is that our empirical beliefs are insulated from our transcendental ones.¹²⁷ Accepting this, Bird applies it to questions about perception, where he finds that two sorts of answers might be given to the question "What do you perceive?" The first will be an answer of the kind "A flash of light," or "An electric discharge," and the second will be of the kind "An appearance," or "A thing in itself." These are clearly very different sorts of answers, and it is not misleading to say that the first is an informative empirical answer and the second an uninformative

¹¹⁸ "Kants Widerlegung des Idealismus" (1958-9).

¹¹⁹ His most recent studies are "Erscheinungsstufung und Realitätsproblem in Kants Opus Postumum" (1953-4); "System und Geschichte in Kants Philosophie" (1958); "Zur Problemanalyse von Kants Nachlasswerk" (1961); and "Zur Frage der Spätereentwicklung Kants" (1963).

¹²⁰ Particularly significant are Schrader, "Kant's Presumed Repudiation of the 'Moral Argument' in the Opus Postumum" (1951); Hübner, "Leib und Erfahrung in Kants Opus Postumum" (1953); Albrecht, "Die sogenannte neue Deduktion in Kants Opus Postumum" (1954); Mathieu, *La philosophie transcendente et l'Opus Postumum de Kant* (1958); and Kaulbach, "Leibbewusstsein und Welterfahrung beim frühen und späten Kant" (1963).

¹²¹ "Kants Widerlegung des materialen Idealismus" (1964).

¹²² B275.

¹²³ *Op. cit.*, pp. 71-74.

¹²⁴ Ak. XVIII. 306 ff.

¹²⁵ *Op. cit.*, pp. 81-82.

¹²⁶ *Kant's Theory of Knowledge* (1962), p. 16.

¹²⁷ *Ibid.*, p. 39.

transcendental answer. The transcendental answer imposes no limits on our ordinary empirical descriptions, for it does not in any way discriminate between things, or even, in any ordinary sense, sorts of things.

It is the failure of the transcendental answer to discriminate between the items perceived that, in Bird's opinion, poses the problem that the Analytic is designed to solve:

by showing how the categories and their associated principles enable us to discriminate between the phenomenal objects in certain allegedly basic ways.¹²⁸

Using this concept of discrimination, Bird is able to maintain some interesting theses. It enables him, for example, to harden his objection to the phenomenalist interpretation of Kant, since, he says, Kant's construction of physical objects is now seen to be "not 'vertical', from low level to higher level descriptions, but 'horizontal', from an indiscriminate manifold of sense to discriminated items within it."¹²⁹ It further enables him to give an interpretation of the distinction between sensibility and understanding as the distinction between the indeterminate manifold (appearance) and the discriminated categorized manifold (phenomenon). The revisionary nature of the interpretation is evident when, on this basis, he can further say that the distinction between the faculties is not strictly required, since:

it is an indisputable truth that we are able to discriminate between items in our perception, and this is enough by itself to introduce a general problem about the ways in which we are enabled to do this.¹³⁰

The second part of the book deals with the explanation of how Kant can say that (1) appearances are necessarily related to the categories, and also (2) objects can appear to us without their being under the necessity of being related to the categories.¹³¹ This is the problem of showing how and why discriminations are necessary. In answering it Bird takes us through the two Deductions and the Analogies focusing his interpretation on the Second Analogy. Here his position is unfortunately unclear. He is certainly right in saying that:

Kant is not interested in the inference from 'The event A-B appeared to take place', to 'The event A-B really took place', but in the prior inference from 'I perceived A and then perceived B' to 'I perceived the event A-B'.¹³²

The point at which there is difficulty is where Bird says that Kant was "right to say that the issue between himself and Hume over [causality] was not that of its usefulness or indispensability,"¹³³ and also that Kant attempted to reinstate the distinction between objective claims and subjective associations of ideas, which is left closed at the end of the Deduction:

by showing that the concept 'cause' was important. His additional step beyond Hume was to show what kind of importance, and what kind of necessity, can be said to belong to this concept.¹³⁴

And what kind of importance is it? It is an importance that is a:

reflection of the fundamental part which this concept plays in our experience, for without it the discrimination of an event, and of an objective time order, would not be possible. The central position of this law explains the importance of the concept 'cause' in our experience in a way in which Hume did not explain it.¹³⁵

But is this showing of the "central position" of the concept any more than showing its "usefulness or indispensability"? And is it not true that Hume did much the same job in the section of the *Treatise* called "Scepticism with regard to the senses."¹³⁶

I would be inclined to take these as merely *ad hominem* objections because I do not think that Kant can do anything else except establish the concept of causality as conditionally necessary. The point has been well made by Bennett, who says that we must distinguish the claims (1) that Kant shows in the Second Analogy that it is necessary that there should be causal laws, and (2) that Kant shows that it is necessary that there should be necessary causal laws.¹³⁷ To adopt (1) would be to adopt the view that Kant charges Hume with underestimating the indispensability of causal laws, while to adopt

¹²⁸ *Ibid.*, p. 43.

¹²⁹ *Ibid.*, p. 57.

¹³¹ *Ibid.*, pp. 57-61; cf. Wolff, *Kant's Theory of Mental Activity* (1963), pp. 157-159.

¹³² *Ibid.*, p. 157.

¹³³ *Ibid.*, p. 164.

¹³⁴ Cf. H. H. Price, *Hume's Theory of the External World* (London, 1940).

¹³⁷ "The Status of Determinism" (1963).

¹³⁰ *Ibid.*, p. 63.

¹³⁴ *Ibid.*, p. 165.

¹³⁵ *Ibid.*, p. 166.

(2) would be to adopt the view that Kant charges Hume with providing a wrong analysis of causal laws. Bennett does not, in this article, attempt to establish (1) for this we must wait for the publication of his book, in the near future, under the probable title of *Kant's Analytic*. He is concerned only with the question of whether there is another conflict between Hume and Kant, i.e., the conflict of (2). Bennett points out that we cannot show the wrongness of the empiricist program for the analysis of causal laws, for:

if we know perfectly well that this whistle's blowing does not cause those workers to down tools, the empiricist has only to ask how we know this, and to amend his analysis in the light of the answer.¹³⁸

What sort of objection could be brought against Hume? Bennett argues that:

just because he thinks that the concept of causality can be analysed in purely empirical terms—and crucially in terms of regularity—Hume does not and cannot attach a fundamental importance to the difference between a rule which holds always and one which nearly always holds.¹³⁹

There could therefore be, and Bennett argues that there is, a dispute between Kant and Hume over whether the body of causal rules shown to be necessary by (1) has to consist of strongly quantified laws ("for all values of . . .") or whether it might consist of weakly quantified laws ("for most values of . . ."). It is further argued that Kant did not, because he could not, establish the former, since:

a preparedness to accept a weakly quantified science is not only permissible but is mandatory upon any scientist who wishes to be able to cope sensibly with a really well attested but unrepeatable experiment, if one should occur.¹⁴⁰

This article should help toward clearing the air, but, conceding the validity of the argument, the question remains of what the statement or statements are for which the concept of causality is conditionally necessary. Wolff has written a book about this question, claiming that Kant will not have answered Hume unless he manages to deduce

the law of causation from the fact of self-consciousness. Such a question Wolff makes appear peculiarly his own by sometimes illegitimate moves. There is a passage from Beck that we may paraphrase as:

The justification of the principles is not merely that they imply the kind of knowledge that Hume doubted, but also that they are implied by "There is a distinction between what is objective and what is subjective."¹⁴¹

This could provide an *ad hominem* argument of enormous force against Hume, and is, in fact, what Wolff himself comes to argue. It is a step in the "strict" deduction of causality if the further requirement of showing that such a distinction is entailed by the fact of self-consciousness can be made out. The passage is, however, construed by Wolff to say that:

The justification of the principles is not merely that they imply the kind of knowledge that Hume doubted, but also that they imply "There is a distinction between what is objective and what is subjective."¹⁴²

This could not be a step in the proof, but at most a corollary of it: it is not, however, what Beck maintains.

The sort of interpretation against which Wolff is arguing can be seen in one of the most stimulating of recent treatments of the Deduction, which comes from Janoska,¹⁴³ who builds his argument around a supposed discovery of an equivocation in Kant's argument made by Aebi.¹⁴⁴ She gave the argument of the Deduction in the following syllogism:

- (1) Each (objective) unity of apperception is a unity according to a rule.
- (2) What makes possible the givenness of a manifold is a (transcendental) unity of apperception.
- (3) What makes possible the givenness of a manifold is a unity according to a rule.

Janoska discusses two possible types of equivocation, employing the distinctions (1) between "subjective₁," or not necessarily true, and "subjective₂," or stemming from the subject, and (2) between "unity of apperception₁" as the unity of what is perceived, and "unity of apperception₂" as the

¹³⁸ *Ibid.*, p. 109.

¹³⁹ *Ibid.*, p. 113.

¹⁴³ *Kant's Theory of Mental Activity* (1963), pp. 48-49.

¹⁴⁴ "Der transzendente Gegenstand" (1954-5).

¹⁴⁴ *Kants Begründung der "Deutschen Philosophie"* (1947). References tying the text to Janoska's argument can be found in Janoska.

¹⁴⁰ *Ibid.*, p. 119.

¹⁴¹ Quoted in Wolff, p. 48.

unity of the act of perceiving an object. It is first argued that if the unity of apperception is to be described as "transcendental," then it must also be described as "objective₁," and this is both consistent with its being described as "subjective₂," and with its not being a source of equivocation. It is then argued that if "unity of apperception" is not to be a source of equivocation, then "unity of apperception₂" must be shown to entail "unity of apperception₁," so that the move can properly be made from the apodosis of (2) through the protasis of (1) to the conclusion (3): that is, "we can only oppose Aebi's interpretation . . . if the unity of the object is effected through the spontaneity of the transcendental subject."¹⁴⁵ Janoska then argues that the move from apperception₂ to apperception₁ is made by Kant in the section of the first edition Deduction where Kant speaks of the transcendental object. The formal argument which Janoska discovers in this section, and which he sets out attractively, is described in the following way:

Under the presupposition that there are, in the Kantian sense, synthetic *a priori* judgments, it follows from the Kantian definition of knowledge that apperception₂ is a necessary condition of apperception₁.¹⁴⁶

This, certainly, is to make Kant's argument a regressive one, i.e., one that fits the pattern of the *Prolegomena* rather than the pattern that we should expect to find in the *Critique*. This does not make it bad exegesis (in fact I think it correct), but in the Deduction as a whole we should expect an argument that goes in the opposite direction.

This Wolff attempts to give us. The book is written as a commentary, which is occasionally irritating, since the only parts of the *Critique* that really interest Wolff are the Subjective Deduction and the Second Analogy (pp. 100-134, 154-164, 260-280). The Objective Deduction and the second edition Deduction have productive imagination excised, and the other Principles are given up for lost.

Wolff divides the Subjective Deduction into four parts (A104-110, 95-97, 97-104, 110-114) which are related as developing stages of one argument. This is an argument which, in the first two stages, operates from the two premisses:

(1) "All the contents of my consciousness are bound up in a unity," which is to be explained

through the fact that "representations (for instance the single words of a verse) distributed among different beings, never make up a whole thought (a verse)."¹⁴⁷

(2) "The contents of my consciousness have the double nature of representations," which is to say that we must distinguish "perception as an object of consciousness" and "perception as consciousness of an object."

With these two premisses the argument of the first stages moves to the conclusion that we have valid synthetic *a priori* judgments through the important step:

(3) "The only way to unify a diversity of mental contents is by referring them, *qua* representations, to an object as ground of their unity."

It is with the step from the second to the third stage that the most novel part of Wolff's account is introduced: an analysis of rule-directed activities to show that they may be spoken of as non-arbitrary, unified, and as correct or incorrect, and the consequent description of "synthesis" as "rule directed reproduction in imagination."¹⁴⁸ In this way steps (2) and (3) are collapsed into:

(2¹) "The only way to introduce synthetic unity into a manifold of contents of consciousness is by reproducing it in imagination according to a rule."

This move is the heart of the Deduction, for, by losing the second premiss, we move from "correspondence" to "coherence"; and:

to say that mental content *R* represents object *Q* is to say that *R* is one of a variety (=manifold) of mental contents which has been, or can be, reproduced in imagination according to the rule which is the concept of *O*.¹⁴⁹

The rest practically writes itself. The fourth stage is a preliminary refutation of Hume, where the categories, as *a priori* rules of synthesis, are held to be necessarily applicable to the contents of consciousness. The fifth stage, contributed by the Second Analogy, completes the refutation with the aid of a subsidiary premiss to the effect that the form of inner sense is time, by showing how the distinction between subjective association and objective con-

¹⁴⁵ *Op. cit.*, p. 203.

¹⁴⁶ *Ibid.*, p. 208.

¹⁴⁷ A352.

¹⁴⁸ *Op. cit.*, pp. 121-125.

¹⁴⁹ *Ibid.*, pp. 133-144.

nection (removed by the Deduction) is made possible through the concept of causality.

The strength of Wolff's book lies in his grasp of the outline of Kant's argument, and in his bringing together with this the insights that we have gained through the long discussions of the private language argument. Wolff could have made this very much clearer if he had continually observed the distinction between general and particular statements—a distinction which has to be considered even in our account of synthesis. When Kant, for example, speaks of synthesis as "putting different representations together,"¹⁵⁰ we can take him to mean either (and perhaps both) of "putting numerically and qualitatively different representations together" and "putting numerically but not qualitatively different representations together." In the first case we shall speak of the apprehension of a particular, and in the second of the apprehension of an objective similarity or universal. Wolff does not make these alternatives clear, and consequently there are suspicious looking movements like the one in the following passage:

To say that A really preceded B is to deny that their order can be changed, now that it has occurred. The real is precisely what we cannot tear up or rewrite. If the order cannot be altered, then it *must* be represented in that way and no other. In other words, we must *always* so represent it. Thus the objective reality of a temporal succession of A and B is expressed by a necessary and universal rule for their representation.¹⁵¹

There is here a half concealed shift, which, if it can be made at all, ought surely to be made proudly and openly, from "If A occurred before B, then A always and eternally will have occurred before B" to "If A occurred before B, then A's always and eternally will occur before B's." Wolff certainly thinks that Kant got from the first to the second, for he says that one of Hume's challenges is "to prove that the observable associations of events are invariable and universal, and hence constitute a sound basis for inferring the future from the past," and also that it is fair to say that Kant met this challenge.¹⁵² I would agree with Wolff that Kant came very close to proving that the first of these statements would not be true unless the second were also true, although not so close as to

remove all problems over induction. This may be seen minimally in that if the argument from Bennett that we have considered is correct, then such a complete proof would be a proof that it is necessary that there should be necessary, i.e., strictly universal, causal laws. It is regrettable that Wolff should obscure so central a point in his rehearsal of Kant's reply to Hume through failure to mark so elementary a detail.

Another objection has frequently been brought against Wolff's book; namely, that his distinction between rules and rules about rules—which corresponds to the distinction between empirical and pure concepts—is insufficiently worked out. This objection is less fair because Kant's own account of the distinction between empirical and pure concepts is so unclear, as Schrader has argued. In an article which is as stimulating as his articles always are, Schrader points out that what is common to both forms of concept is (1) that they are rules of combination, and (2) that they originate in the understanding (obtain there the "form of generality"). Kant's attempt to link empirical concepts with abstraction from experience must be rejected as un-Critical, for:

the essential difference between empirical and *a priori* concepts in the *Critique* is not that the former are abstracted from intuition whereas the latter are contributed by the understanding, but rather that the former are contingent while the latter are necessary.¹⁵³

Will this do as a distinction? Schrader argues that the *a priori* concepts are "contingently necessary" in that they are "necessary as conditions for the possibility of experience and contingent in that they are valid only for possible experience,"¹⁵⁴ and yet the same thing may also be said of the empirical concepts. The difference is not a difference in kind, but involves a continuum of changing degrees of universality. So far I think that Schrader is correct: it is only by looking at Kant in his historical context that we shall find explanation of why Kant continued to maintain contrary views.

§7. SYNTHETIC AND *A Priori*

The *Critique* was written to explain the possibility of synthetic and *a priori* judgments. For many

¹⁵⁰ A77=B103; cf. Wolff, pp. 68–70.

¹⁵¹ *Ibid.*, pp. 266–267.

¹⁵² *Ibid.*, p. 163.

¹⁵³ "Kant's Theory of Concepts" (1957–8), pp. 270–271.

¹⁵⁴ *Ibid.*, p. 273

people today this provides an insuperable barrier to their introduction to the *Critique* because, in their opinion, there are no synthetic and *a priori* judgments. The subject has now become so complex that I cannot here do more than indicate the outlines of the controversy and try to point out the route that will take us back to the Kantian concept.

We may begin with a controversy provoked by Robinson in an article on the Kantian use of the word "necessary." Four senses were distinguished: (1) the "Aristotelian sense" in which a necessary proposition is one containing the word "must" or some cognate expression, (2) the "compulsory belief sense" in which a necessary proposition is one which, for one reason or another, it is "necessary for us men to believe," (3) the "Leibnizian sense" involving logical consistency, and (4) the "universal sense" according to which a necessary proposition is one "which asserts a universal connection with unrestricted universality."¹⁵⁵ Robinson then goes on to argue that:

Kant's concept of a necessary proposition is nothing definite, but just a confusion of the four clear concepts of a necessary proposition which I have indicated.¹⁵⁶

He also argues that because Kant departs from the Leibnizian sense he has a problem about the connection of the concepts of necessity and truth, but this need not concern us greatly, as Kant was himself aware of the problem, and showed this in asking about the possibility of synthetic and *a priori* judgments.

About the first two senses little need be said. Although Robinson argues for the Aristotelian sense as a distinct type of necessary proposition,¹⁵⁷ Bird argues cogently that it is not.¹⁵⁸ He also treats nicely of the compulsory belief sense, admitting that Kant does have a compulsory belief sense of the word, but denying that Kant was confused about this.¹⁵⁹ There is, for all that, an element of compulsory belief present more generally in necessity. It is not that a compulsive belief provides a basis for calling a proposition necessary, as Parkinson would appear to suggest,¹⁶⁰ but rather that the negation of a necessary proposition is either in-

consistent or involves some form of absurdity that would preclude its being the object of a rational belief.

The Leibnizian sense need not concern us for the moment. It is certainly true that Kant did not think of "necessary proposition" and "proposition whose negation is logically inconsistent" as being identical in meaning, even though he thought that any proposition whose negation was logically inconsistent would be a necessary proposition.

The universal sense is introduced on the basis of Kant's claim that we have *a priori* knowledge either (1) if we have a proposition which in being thought is thought as necessary, or (2) if we have a proposition which is thought in "strict universality," i.e., in such a way that "no exception is allowed as possible."¹⁶¹ Kant is quite clear that a proposition is thought as necessary if and only if it is thought with strict universality, and his reason for this would appear to be the same as that for which we would today say that someone who refuses to admit the possibility of recalcitrance to one of his general propositions is making it analytic. Kant did not say that all unrestrictedly universal propositions were necessary, nor did he say that all true unrestrictedly universal propositions were necessary, but that those unrestrictedly universal propositions were necessary which we would not permit to suffer recalcitrance. Vaihinger speaks of the two criteria as being based upon our concepts of a qualitative and a quantitative *Anders-sein-können*, and we might, alternatively, speak of them as the intensional and the extensional aspects of necessity. We certainly cannot speak of them as two independent and fortuitously coincident specifications of the *a priori*.

Kant did not therefore think of strictly universal propositions as a variety of necessary proposition, but he thought of strict universality as a feature of necessary propositions. We can avoid difficulties of intensions by putting the main weight upon the second of Kant's two criteria for *a priori*, treating strict universality as a means for determining when an intensional connection has been made.

The problem is now posed that the extensional consequence of necessity that we are now using as a test of necessity is today part of a behavioral test

¹⁵⁵ "Necessary Propositions" (1958), p. 291.

¹⁵⁶ *Ibid.*, p. 293.

¹⁵⁷ *Ibid.*, pp. 290-291.

¹⁵⁸ "The Necessity of Kant" (1959), pp. 389-390.

¹⁵⁹ Cf. A824-9=B852-7.

¹⁶⁰ "Necessary Propositions and *A Priori* Knowledge in Kant" (1960), p. 394.

¹⁶¹ B3-4.

that we often apply in finding out whether someone is maintaining a statement as analytic. If this is the case, then the class of synthetic and necessary propositions is of necessity an empty one. The belief that the class is an empty one had been traced by Beck to a change of meaning that has occurred in the word "analytic" since Kant's day. His nicely modelled argument claims that as the class of synthetic and *a priori* truths¹⁶² came to be regarded during this century as an empty one, the word "necessary" suffered a change of meaning so that it is now equivalent to "analytic" rather than, as formerly, to "*a priori*." If this is the case, then we should expect that if the older sense of "analytic" (analytic₁) and "*a priori*" did have distinct roles, there will be some equivocation in the contemporary use of "analytic" (analytic₂). Do we in fact find this?

Let us take a theory of the analytic which holds that a statement is analytic to the extent or degree to which it will be held impregnable against revision by experience. . . . Then there are two ways in which a proposition may be found to be analytic (i.e., analytic₂): (1) by inspection of the sentence itself, if it is logically or linguistically true; and (2) by investigation of its role in an organised body of experience we call knowledge.¹⁶³

This "theory of the analytic" is, I would argue, the same as the extensional aspect of the Kantian theory of the *a priori*, and the two tests, which Beck calls the microscopic and the macroscopic tests, are, as he argues, exactly correspondent to the two bases for the attribution of necessity that distinguished analytic₂ from the synthetic *a priori*. It seems clear in outline, however unclear it may be in detail, that some such distinction is operative between the sorts of justification that someone might present for not permitting the possibility of a falsifying instance. But if someone is going to maintain a statement as necessarily true and is not going to maintain this on purely definitional grounds, then he will in effect have to answer the

Kantian question as to the possibility of synthetic *a priori* judgments.

This sort of defense of the Kantian category of the synthetic *a priori* is one that is today being increasingly pressed. Hanson insists that there is a distinction between "characterizing the *structure* of propositions" and "characterizing the *mode of justification* of propositions,"¹⁶⁴ and Stenius holds, what is at base equivalent, that the distinction between analytic and synthetic is "a semantical distinction," while that between *a priori* and *a posteriori* is "an epistemological distinction."¹⁶⁵ There are still underlying difficulties that can best be uncovered by concentrating on the concept of definition.¹⁶⁶ The argument has been that there is a difference between (1) being analytic as being held to be unfalsifiable, which is the Kantian *a priori*, and (2) being analytic as being reducible to a truth of logic by means of definitions, which is fairly close to the Kantian analytic. While there has been at times a tendency to confuse the two, some have noticed the gap and closed it with a concept of being complicatedly analytic, i.e., while seeing that there are many statements which are apparently unfalsifiable, and also seeing that it would be implausible to suggest some definitional equation as the basis of this unfalsifiability, vague appeal is made to some more recondite form of definition that will take us from one to the other. In this way an amorphous concept of definition is allowed to occupy all the territory that was to be held by Beck's macroscopic test for being analytic.

Concentrating on the concept of definition can help us to concentrate the dispute, because most of the articles dealing with the distinction between analytic and synthetic propositions can be interpreted in terms of it. The important contribution of Waismann's classic articles has come to be seen in their demonstration that the meaning of a word cannot always be given (and cannot perhaps ever be given) through a definitional equation.¹⁶⁷ Yet it is such an equation that is required as a substitution license in the reduction of non-logical to

¹⁶² I speak alternatively of "synthetic and *a priori*" or "synthetic and necessary," accepting Kant's equivalence of "*a priori*" and "necessary." This connection seems unproblematic.

¹⁶³ "The Meta-Semantics of the Problem of the Synthetic *A Priori*" (1957), p. 231.

¹⁶⁴ "The Very Idea of a Synthetic-*a priori*" (1962), p. 521.

¹⁶⁵ "Are True Numerical Statements Analytic or Synthetic?" (1965), p. 359. Other valuable treatments of the special question over the status of geometrical truth are to be found in Martin and Körner, and in Locke, "Mathematical Statements," *Australasian Journal of Philosophy*, vol. 41 (1963), pp. 186-197, and J. Hintikka, "Are Logical Truths Analytic?" *The Philosophical Review*, vol. 74 (1965), pp. 178-203.

¹⁶⁶ The Kantian concept is well handled in Beck, "Kant's Theory of Definition" (1956). Cf. Crawford, "Kant's Theory of Philosophical Proof" (1961-2), and Beck, "Can Kant's Synthetic Judgements Be Made Analytic?" (1955-6).

¹⁶⁷ "Analytic-Synthetic" (1950).

logical truths. Weitz, developing the view, claims that there are all sorts of statements "whose analytic character has nothing to do with their reducibility to logical truths,"¹⁶⁸ and this resistance to the definitional account is produced by his recognizing that analytic character depends on meanings, and meanings cannot be measured by definitions as these are ordinarily understood. If, however, we extend the concept of definition, we can reallocate the analytic character of these statements in their definitional role. Thus H. G. Alexander says that "one might attempt to distinguish different types of necessary statement . . . by considering the different methods of defining the constituent words,"¹⁶⁹ and then, distinguishing three main means of definition, he points out that the notion of substituting *definiens* for *definiendum* makes sense only in the case of verbal definition, and not in the cases of ostensive definition or definition in use.

Faced with this difficulty we must either choose to say that there are some necessary truths which are not analytic by definition, or to say that "analytic" and "necessary" are coextensional terms, but not all analytic truths are reducible to truth of logic by means of definitions. A strong case can be made for adopting the first of these alternatives, although to adopt it is not to accept a sharp line between analytic and necessary propositions. This, of course, must go.

If our previous argument can be sustained, then we have indicated a class of propositions which are necessary but which are not analytic, and this class will probably contain such a favored example as "Nothing can be red and green all over at the one time" as well as "Every happening has a cause." The question arises of how we are to delimit within this class those propositions that were of particular concern to Kant.

The question may be taken together with another attack that has been launched upon the notion of an analytic proposition, and which has been taken to introduce a blurring of the line between analytic and synthetic propositions rather than the line between analytic and necessary propositions. This is the attack made by Quine,¹⁷⁰ and which may be

broken into two parts. The first part may be represented as accepting the account of analytic propositions as those reducible to truths of logic by means of definitions, and then stipulating that the definitions involved will either be stipulative or lexical, and so concluding that analytic propositions will either be trivial or their analytic nature will founder on one idiosyncratic idiolect. Quine does not find problematic those analytic propositions which are based upon explicit stipulative definitions ("nominal synthetic definitions" in the Kantian scheme), and we may accept these as the paradigm of analytic propositions. We may extend the notion to include analytic propositions which are so because of implicit stipulative definitions by saying that those propositions which a given person holds as necessary, and for which he will accept a suggested stipulation as the basis of their necessary truth, are to be taken as analytic. Such an extension is not an easy one to make, as White¹⁷¹ and Goodman¹⁷² as well as Quine make clear, but I think that it can be made, and, if so, we shall still have a class of propositions which are not analytic but which are necessary.

In the second part of his attack Quine takes "analytic" to be equivalent to the sense of "necessary" that we have taken to be the Kantian usage. He here argues against the idea that we can speak of analytic or synthetic statements on the grounds that the former are "verbally" rather than "factually" true. If we should meet recalcitrance, then some change must be made in our conceptual scheme, but this change can be made at more than one point; and, in theory, there is no statement which is theoretically immune from revision. We may, he allows, draw distinctions between statements on the grounds of their relative immunity from revision, but this will be to make the distinction between analytic and synthetic statements one of degree and not of kind.

This view has been variously and continuously attacked—particularly good treatments being those by Grice and Strawson¹⁷³ and by Bennett¹⁷⁴—but from a Kantian point of view it is wholly acceptable once we have remarked that Quine is not talking about the class of analytic, but about the

¹⁶⁸ "Analytic Statements" (1954), p. 490.

¹⁶⁹ "Necessary Truth" (1957), p. 517.

¹⁷⁰ "Two Dogmas of Empiricism" (1951).

¹⁷¹ "The Analytic and the Synthetic: an Untenable Dualism" (1950).

¹⁷² "On Likeness of Meaning" (1949).

¹⁷³ "In Defence of a Dogma" (1956).

¹⁷⁴ "The Analytic and the Synthetic" (1958-9).

class of necessary propositions. Of course Kant was inclined, in his talk of the separate contributions of sensibility and understanding, to think of a verbal and a factual component entering into the meaning of any given proposition, but this, with all its corollaries, has long been an object of distaste to commentators. There will no longer be a sharply delimited class of synthetic and *a priori* propositions, nor, if Quine is right, will there be any propositions which are totally immune from revision in the face of recalcitrant experience. We retain, however, the essential Kantian insight that there are propositions which are held to be necessary, i.e., strictly universal, because they are necessary for some, or all, conceptual activity.¹⁷⁵

§8. THE CATEGORICAL IMPERATIVE

It would be pointless to review here Beck's *Commentary on Kant's Critique of Practical Reason* (1960). His name is well known as that of a scholar possessed of philosophical intelligence, and both qualities are markedly present in the book. It will remain as indispensable to students of Kant's ethics as are the works of Vaihinger, Kemp-Smith, and Paton to students of the first *Critique*.

More to the point is Gregor's study of the *Metaphysic of Morals*,¹⁷⁶ which is a reflection both of the unflagging interest in the question of the application of the categorical imperative, and in the tendency to study Kant's ethics as much through his later writings as through the *Foundations*. Paulsen's conclusion that the *Metaphysic of Morals* is "freilich geringwertig" is now a contested one, although contested on various grounds. Brown argues that its first part "is interesting and important for its exposure of a fatal defect in Kant's moral theory, and for no other reason"¹⁷⁷—this defect being that Kant's handling of casuistical questions shows that "he has no principles other than formal consistency on the one hand and *ad hoc* convictions on the other."¹⁷⁸ The *Metaphysic of Morals*, on this view, is important for showing the impossibility of deriving any concrete obligations from the categorical imperative, and Brown, in

accordance with this view, is inclined to give credence to Duncan's interpretation of the *Foundations*, where he opposes "ethical," "metaphysical," and "critical" interpretations, and argues that the categorical imperative is not a moral criterion, but merely a description of an agent's motive when he is acting in a morally good manner.¹⁷⁹ We may concede Duncan's contention that the *Foundations* is intended to further Kant's Critical scheme,¹⁸⁰ but the idea seems scarcely tenable that Kant did not intend the categorical imperative as a formula, that is, as a recipe for constructing particular categorical imperatives.¹⁸¹ That, as a formula, it is inadequate without extra *ad hoc* assumptions is not ground enough for denying Kant's intention to use it as such. That such extra *ad hoc* assumptions are involved in the application of the categorical imperative is argued interestingly, and in a less *ad hominem* way than by Brown, by Haezrahi, who claims that:

an action which accords with the categorical imperative, even if performed for the sake of the law defined through the imperative, is not necessarily a morally right action.¹⁸²

This is the claim that some impermissible acts are licensed by the categorical imperative, and to substantiate it Haezrahi argues that somebody who acts on the maxim "Establish a world dictatorship for the sake of the ideal of absolute power" has a will which is (1) consistent, (2) free, (3) autonomous, (4) of universal validity, and (5) capable of imposing a definite, rationally consistent order on the world. She argues that what supplies the categorical imperative with moral content (and eliminates this maxim) is given by the:

proposition that in the moral domain all members of the human race enjoy equality of status and hence possess a certain intrinsic worth *qua* human beings.¹⁸³

In a later article Kant's proof of this proposition is examined and shown to be defective, and, while its importance as the real as well as formal cause of moral experience is insisted on, it is shown that "an

¹⁷⁵ Cf. Hamlyn, "On Necessary Truth" (1961).

¹⁷⁶ *Laws of Freedom* (1963).

¹⁷⁷ "Has Kant a Philosophy of Law?" (1962), p. 48.

¹⁷⁸ *Ibid.*, p. 44.

¹⁷⁹ *Practical Reason and Morality* (1957), pp. 53-56 etc.

¹⁸⁰ Cf. Silber, "The Metaphysical Importance of the Highest Good etc." (1959).

¹⁸¹ Cf. Paton, "The Aim and Structure of Kant's *Grundlegung*" (1958).

¹⁸² "The Avowed and Unavowed Sources of Kant's Theory of Ethics" (1952), p. 159.

¹⁸³ *Ibid.*, p. 167.

act of faith, and a gratuitous act of faith at that, is needed for its acceptance."¹⁸⁴

These two articles are stimulating and, in the main, correct. My only difficulty is with Haezrahi's conception of consistent willing, and, this because the ideal of the dignity of man is built into Kant's test of consistent willing. Yet it is not built into it as directly as Gregor, who also allows that there is a need for criteria if we are to derive particular categorical imperatives from the formula of all categorical imperatives, makes out when she says that we must:

analyse the nature of man as a composite, animal, rational, and moral being, and compare the maxim in question with the order which reason finds among these elements in his being.¹⁸⁵

We then, she believes, discover the connection between being free and being rational, and being rational becomes acting in accordance with the formula "Act only on that maxim which, if it became universal law, would advance systematic harmony amongst human beings." Any maxim which does not fit this formula is one which we, as rational beings, "cannot consistently will."

This idea of a systematic harmony must find its place in Kantian exegesis,¹⁸⁶ and yet on the strict Kantian view it ought to be derivable from the formula of universal law. This presupposes that this formula can be made workable as a moral criterion. That it can be made workable has been argued recently by several authors, as Ebbinghaus, who believes it a mistake to say that Kant's ethical theory "confines moral philosophy to stating what the concept of duty is simply as regards its form, and makes impossible the articulation of particular duties that are materially different,"¹⁸⁷ or Singer, who mentions and disagrees with the idea "that the categorical imperative is incapable of being used to establish any moral rule or to settle any concrete question of ethics."¹⁸⁸

A number of examples have been debated in an attempt to show how "being able to be consis-

tently willed" functions as a moral criterion. Not the least important of these is the lying-to-the-murderer case, where Singer argues that Kant misapplies his own criterion, by slipping from a sense of "unconditional" meaning "not conditional upon the desires of the agent" to one meaning "not conditional upon anything."¹⁸⁹ The nature of the distinction is further clarified by Beck.¹⁹⁰ Another interesting example is considered by Ebbinghaus, who argues that a person cannot subject himself to an arbitrary power since "it is self-contradictory that a will should be able to will its own annulment with the necessity of law in every possible exercise of its own volition."¹⁹¹ Further examples can be found in Singer and Gregor, but, naturally, attention has tended to be focused upon Kant's own four examples in the *Foundations*:

"Always to shorten life if its continuance threatens more evil than it promises pleasure"; "Always to borrow money with promises to return it, when I am in need, even though I know that I shall not do so"; "Always neglect natural gifts"; and, "Always refuse to aid others in distress."

One of the two operative distinctions determining the list is that between perfect and imperfect duties, which has been usefully discussed by Gregor, who shows the development of Kant's views from the time of the *Vorlesung* to the *Metaphysics of Morals*.¹⁹² She shows how the distinction was at first that between types of obligation, the legal and the ethical, but that in the *Foundations* Kant recognized perfect duties to oneself, so that here he understands by a perfect duty "one which allows no exceptions in the interest of inclination,"¹⁹³ and so creates a distinction between strict and wide obligation. Gregor argues that construing imperfect duties as those which do permit arbitrary exceptions is consistent with what Kant says, even though there are some passages in the later writings which favor a rigoristic interpretation.

This seems to me to be a reliable version of the distinction, and yet it is difficult to see how it

¹⁸⁴ "The Concept of Man as an End in Himself" (1961-2), p. 223.

¹⁸⁵ *Laws of Freedom* (1963), p. 205.

¹⁸⁶ Cf. Paton, *The Categorical Imperative* (London, 1947), p. 150; Beck, *Commentary . . .* (1960), p. 99; Singer, *Generalisation in Ethics* (1961), pp. 261-263.

¹⁸⁷ "Interpretation and Misinterpretation of the Categorical Imperative" (1954), p. 99.

¹⁸⁸ *Op. cit.*, p. 218.

¹⁸⁹ "The Categorical Imperative" (1954); cf. Paton, "An Alleged Right to Lie" (1953-4), and Ebbinghaus, "Kants Ableitung des Verbotes der Lüge aus dem Rechte der Menschheit" (1954).

¹⁹⁰ "Apodictic Imperatives" (1957); cf. Moritz, *Kants Einteilung der Imperative* (1960).

¹⁹¹ *Op. cit.*, p. 103.

¹⁹² *Op. cit.*, pp. 95-112.

¹⁹³ Ak. IV. 421 ff.

relates to the "two criteria" that Kant gives of (1) maxims which cannot be thought as universal laws of nature without contradiction, and (2) maxims which cannot be willed as universal laws of nature without inconsistency,¹⁸⁴ because of the difficulty of getting clear about what exactly these two criteria are.

Although the cases of suicide and neglect, as perfect and imperfect duties to oneself,¹⁸⁵ are perhaps more interesting from the point of view of getting at the Kantian presuppositions, it is the duties of promise keeping and benevolence, as perfect and imperfect duties to others, that naturally collect most discussions. Some of these I shall now mention.

Ebbinghaus argues that in adopting the maxim of never giving aid to others a person has an inconsistency in his will in that, through the universalization of his maxim, he is willing behavior on the part of others that could cut across his own self-interest, and, as Ebbinghaus believes, "a man cannot in harmony with his own will choose to be abandoned in misfortune by those who could give him help."¹⁸⁶ There is nobody today who argues that Kant is, in this example, introducing an appeal to self-interest, for there is no question of the person having to consider the possible practical consequences of his action in terms of other people following his example. Yet despite these claims, I am not sure that there is not a complexly prudential type of reasoning going on here—something that is half-moral, and yet not quite moral. This becomes more apparent in Harrison, who takes Kant to mean by "inconsistency in the will" in this context that:

though there is a motive for willing the universal adoption of the maxim [not to help others], for if my maxim were universally adopted, I should not have the disagreeable task of helping others in distress, there is also a motive against, for I would in this case not have the agreeable experience of being helped by others.¹⁸⁷

Is it not clear that someone who willed the universal adoption of the contrary maxim always to help others would also, on *this* criterion, have an incon-

sistency in his will? It is just that now what was previously "for" is now "against." In the face of inconsistency in either direction, a man might well choose to be abandoned in misfortune if he thought that his self-interest would best be advanced by a policy of never helping others and never being helped by them. And it is worth noticing that such a person is not willing the abolition of mutual aid. His maxim (in the simplest case) is "If you do not believe that your interest will ever be advanced through the help of others, then do not help others," and this does not entail, as Ebbinghaus believes, that everybody is authorized by the maxim to refuse to help others in so far as he is himself immune from need, but entails only that those who are honestly prepared, the world being as it is, to waive their right of protest, should refuse to give help.¹⁸⁸

In fact, the inconsistency is not that of being unable to give an affirmative answer to "Are you honestly prepared to face the concrete consequences of people not helping you as a result of your maxim?" or to "Are you honestly prepared, in the world as it is, to waive your right of protest?"¹⁸⁹ the first of which would make the inconsistency a simple prudential inconsistency, and the second of which makes it what I have called a complexly prudential inconsistency. Many people would be able to get through these questions who would not be able to give a positive answer to "Are you honestly prepared, in the world as it conceivably might be (i.e., with you stripped of your fortunate position), to waive your right of protest?" There is this second counterfactual condition involved—the condition to which Haezrahi insists we should direct our attention—which turns the inconsistency into that of doing as one would not wish to be done by.

There are certain difficulties in the idea that Kant's criterion of morally impermissible actions as those whose universal adoption cannot be consistently willed is to be construed as forbidding those which would involve violation of the Golden Rule. A first difficulty would be that in making the inconsistency that of licensing others to do things

¹⁸⁴ Ak. IV. 424.

¹⁸⁵ Notes "On Duties to Oneself" by Mavrodes and Narveson, *Analysis*, vol. 24, (1964), pp. 165–168, were provoked by J. Meiland, "Duty and Interest," *Analysis*, vol. 23 (1963), pp. 106–110.

¹⁸⁶ *Op. cit.*, p. 106.

¹⁸⁷ "Kant's Examples of the First Formulation of the Categorical Imperative" (1957), p. 56.

¹⁸⁸ Singer, *op. cit.*, p. 270, argues that this does entail everybody, since nobody could, in fact, be prepared to waive this right of protest.

¹⁸⁹ I take this to be equivalent to: "Would you be prepared to keep the world just the way it is except that now nobody ever helps anybody else?"

which could, under the two counterfactual conditions, cut across my self-interest, the form of the Golden Rule that is being invoked is "Do not do to others what you would not want them to do to you," whereas it is often quite proper to do to others what I would not like being done to myself. I think it is this point, that is also valid against the stronger Golden Rule which is virtually equivalent to what Singer has called the Generalisation Principle,²⁰⁰ that is made in Kolenda's article when she maintains that "the moral worth of the maxim really depends on the worth of the material volition against which it is tested,"²⁰¹ and that where we ought to apply the categorical imperative "is a question which [it] cannot, of itself, supply."²⁰²

The categorical imperative has frequently, as at the back of this discussion, been taken as equivalent to the Generalisation Principle. The same is true of an article by Hall, in which he argues that we must sharply distinguish between the formula of universal law and the formula of the law of nature—the latter involving temporal conditions that are foreign to the former.²⁰³ Underlying his discussion there exists the same assumption. But the Generalisation Principle, as is being increasingly pressed, is merely a tool that we may use to get someone to look at his moral beliefs "from the outside," thus uncovering self-deceit, and so discovering to a person an inconsistency in his beliefs.²⁰⁴ Yet Kant thought that there were maxims which could not be willed as universal law in a stronger sense than this. He thought, particularly in the case of inexpressible duties, that he could convict someone of violating the Generalisation Principle without reference or appeal to his introspective honesty, and this by giving a test of what a rational person could will. Harrison is interesting in this connection, in that he says that Kant's categorical imperative is "not just telling us that what is right for me

is right for any similar person similarly placed," and yet can also interpret the wider criterion to mean that "if it were within our power to bring about a state of affairs in which everyone acted on our maxim, we could not bring ourselves to do it."²⁰⁵ This has the effect of reducing the categorical imperative to the Generalisation Principle, and Kemp quite rightly objects to the account by saying that Kant's wider criterion is "due to, or at least connected with, the fact that it would be irrational (not impossible) to set oneself to act accordingly."²⁰⁶ Harrison rejoins by saying:

you have to know what is immoral first in order to know what a Rational Being would choose, and so cannot deduce that it is immoral from the fact that a Rational Being would not choose it.²⁰⁷

This puts us back with the problem of trying to turn Kant's criteria into a method of convicting someone of inconsistency of the will by reference merely to the fact that he is a rational being, and without reference to any moral principle which, as a matter of fact, he might possess.

An extremely plausible attempt at this has been made by Singer. After quoting Hegel to the effect that "a contradiction must be a contradiction of something, i.e., of some content presupposed from the start as a fixed principle," Singer points out that the maxim of an action, which is what the categorical imperative is designed to test, is itself a determinate principle of conduct which possesses a content, and that Kant never implies that the existence of property, or life, or the institution of promise keeping, is a logical necessity.²⁰⁸ Examination of maxims reveals, however, that "to say that someone is acting on a certain maxim is to imply (if not to say) that he is acting for a given purpose,"²⁰⁹ and we may therefore say that there is an inconsistency in someone's will if his purpose in

²⁰⁰ Cf. discussion of the Golden Rule in §9.

²⁰¹ "Professor Ebbinghaus' Interpretation of the Categorical Imperative" (1955), p. 76.

²⁰² *Ibid.*, p. 77.

²⁰³ "Kant and Ethical Formalism" (1960-1).

²⁰⁴ Recent discussions of self-deceit have not been brought into relation to Kant, but ought to be, if only because someone deceiving himself is someone who is, in some sense, located in the midst of an inconsistent set of propositions. See Demos, "Lying to Oneself," *The Journal of Philosophy*, vol. 43 (1960), pp. 588-595; Canfield and McNally, "Paradoxes of Self-Deception," *Analysis*, vol. 21 (1961), pp. 140-144; Canfield and Gustavson, "Self-deception," *Analysis*, vol. 23 (1963), pp. 32-36; Penelhum, "Pleasure and Falsity," *American Philosophical Quarterly*, vol. 1, (1964), pp. 81-100; Siegler, "Demos on Lying to Oneself," *The Journal of Philosophy*, vol. 54 (1962), pp. 469-478, "Self-deception and Other Deception," *The Journal of Philosophy*, vol. 55 (1963), pp. 759-764, "Self-deception," *Australasian Journal of Philosophy*, vol. 41 (1963), pp. 29-43.

²⁰⁵ *Op. cit.*, p. 52.

²⁰⁶ "Kant's Examples of the Categorical Imperative" (1958), p. 64.

²⁰⁷ "The Categorical Imperative" (1958), p. 360.

²⁰⁸ *Generalisation in Ethics* (1961), p. 252.

²⁰⁹ *Ibid.*, p. 244.

adopting that rule would be defeated if everyone were to adopt it. This account has the valuable consequence of showing that there is not a sharp distinction between the "narrow" and the "wide" criterion, in that the reference to purpose is essential even to cases of the narrow criterion. The maxim, "Make promises with no intention of keeping them in order to further one's own interest," is not one which everyone could act on all the time, but the maxim, "Make promises with no intention of keeping them in order to destroy the purpose of promise keeping," is a maxim upon which everybody could act, although not for very long, without the slightest inconsistency of will. Further, the example suggests the possibility of turning all cases of the wider criterion into cases of the narrower criterion by putting the purposes of the maxim into the maxim itself; and this is something which Singer himself proposes.²¹⁰

I think that Singer is substantially right about all this, and, although much of what he says can be found in other authors, his account is valuable for its organization and for the wealth of examples with which he sustains his argument. The impasse we had previously reached was that a person is describable as having an inconsistency in his will if he adopts a maxim which he does not believe to be a moral law. We are now in a position to remove the reference to his body of moral beliefs by saying that he has an inconsistency of the will if he adopts a maxim (in the extended sense) upon which not everybody can act all the time. Honesty is now only required in ascertaining motive. But if not everybody can act on the maxim, then not everybody ought to act on it, and therefore somebody ought not to act on it, and therefore, finally, no one ought to act on it without some reason or justification. We are also indebted to Singer for provoking a discussion of these moves, and although some outlines are clear, final resolution has not yet been reached.²¹¹

§9. THE GOLDEN RULE

In the previous section I have suggested that someone is spoken of as having an inconsistency in

his will when he violates the Golden Rule, with the qualification that we can know when there is an inconsistency in someone's will by knowing it to be impossible for him consistently to will the maxim of his action, because if universally adopted the purpose to be achieved through the action would be defeated. Clarity may be gained through two main studies: the one by Reiner, who traces the history of the Golden Rule, distinguishes three main forms of it, and discusses their merits and demerits,²¹² and the other by Singer, who covers a lot of the same ground as that in the later part of Reiner's article, but provides a much more detailed defense of it against various of its critics.²¹³

One point of particular Kantian relevance is the relation of the Golden Rule [GR], "Do unto others as you would have them do unto you," to the Generalisation Principle [GP], "If an action is right (wrong) for one person, then it is right (wrong) for any similar person in similar circumstances." Singer has treated both [GR] and [GP], and, although his statements on their relation are equivocal, the outlines are clear.²¹⁴ Attempting to obviate certain difficulties in [GR] he interprets it as "requiring *A* to act towards *B* on the same standard or principle that he would have *B* apply in his treatment of him." A first interpretation of this would be to take it as requiring *A* to act towards *B* on the same standard that he would want *B* to apply in his treatment of him. It is against [GR] in this form that objections have been lodged, since *A* might well, if the roles were reversed, desire *B* to act on a different principle from that which *A* is acting upon, even though he would not condemn *B* for acting on this principle.²¹⁵ This interpretation suffers from one of the defects of the inversion of [GR], "Do unto others as they would have you do unto them," which is inadequate in any form, since it either devolves into "Always meet other people's emotional demands on you," or into "Always act on other people's principles and never on your own." We should act on the inversion of [GR] only when it is reasonable to do so, in the same way that we should act on the principle of doing what we would like to have done to us only when it is reasonable to do so.

²¹⁰ *Ibid.*, pp. 277-278.

²¹¹ The controversy to be reviewed shortly in this journal together with other recent work on ethics.

²¹² "Die Goldene Regel" (1948).

²¹³ "The Golden Rule" (1963).

²¹⁴ In "The Golden Rule," p. 309, he says that [GR] is "the source or at the basis" of [GP], while in *Generalisation in Ethics*, p. 16, he says that [GR] is an "immediate consequence" of [GP].

²¹⁵ Cf. Reiner, *op. cit.*, pp. 84-85; Singer, *op. cit.*, pp. 298-299.

In a second interpretation we may take [GR] as requiring *A* to act on the same standard or principle that he would be prepared to see *B* applying in his treatment of him. Here "would be prepared to see" is equivalent to "would not take moral exception to," and is contrasted with "would not take emotional exception to." Here the connection with [GP] becomes clear, for from the tautology "You ought to do to others what it is right for you to do" we can get, through [GP], to "You ought to do to others what is right for anyone at all to do to anyone at all," and [GR] is derivable from this by instantiation.

Although the first interpretation of [GR] leads to a moral principle which frequently gives counter-intuitive results, Reiner insists that it is not without value. This may be brought out by noticing that in teaching a child we do use arguments like "If you don't like it being done to you, then you ought not to do it to others." This is not a valid argument, but in the full context of the child's situation it is used (1) to show the child that other people's wishes are relevant to his own conduct, (2) to show him when other people's wishes are relevant, and (3) to show him the connection between a moral prescription addressed to himself and his right to protest at other people's treating him with less consideration. This point relates to one of Singer's remarks:²¹⁶

If I would have others take account of my interests and wishes in their treatment of me, even though my interests and wishes may differ considerably from their own, then what the Golden Rule in [the second] interpretation requires of me is that I should take account of the interests and wishes of others in my treatment of them. I am to treat others *as* I would have them treat me, that is, on the same principle or standard as I would have them apply in their treatment of me.

It is here suggested that [GR] on the second interpretation makes two equivalent demands: (1) that I should take account of the interests and wishes of others in my treatment of them, and (2) that I should act toward others on the same principles that I would be prepared to have them apply in their dealings with me. These are clearly different, in that (1) is a moral rule enjoining consideration for others, while (2) is a special form of

[GP]. The two are nevertheless easily confused, and it is worthwhile to notice the reason for this. That someone who violates (2) will generally be someone who violates (1) is a consequence of the extreme generality of the rule enjoining consideration for others. But that someone who violates (1) will generally be someone who violates (2) is based upon the fact that a person who breaks any moral rule will in general also be someone who violates [GP]. This remark is based upon the empirical fact that people are far more prone to make exceptions in their own favor than they are to challenge moral rules.²¹⁷ In view of the fact that Kant often speaks of a person who has inconsistency in his will as of one who is making an exception in his own favor, it is worth insisting that "to make exceptions in one's own favor" and "to violate [GP]" are equivalent expressions. This is necessary because of a puzzling treatment of the phrase by Monroe,²¹⁸ who holds that a person who tries to prevent or who does not want others to act on a principle that he would be morally prepared for them to act on is making an exception in his own favor. This seems to me a misuse of the expression, since the following statements are true: (1) a person who does not do what he allows that he ought to do is not making an exception in his own favor, but is simply doing something wrong, (2) a person who does not want to do what he allows he ought to do is, by this fact alone, neither making an exception in his own favor nor doing anything wrong, and (3) a person who prevents others from acting as he admits they ought to act may, depending upon his moral beliefs, either be making an exception in his own favor, or be doing something wrong, or be doing something right. The foundation for these statements lies in the logic of exceptions. If I am doing *X*, then I found an exception to the statement "nobody is *X*-ing," but I do not necessarily found an exception to "nobody ought to be *X*-ing." It is a mistake to suppose that someone who excepts himself from the scope of an obligation, in the sense of not acting upon it, is therefore making an exception in his own favor. If he recognizes and concedes his culpability, then he is not adding an "except"-clause to any universal proposition. People have confused going against one's principles with making an exception in one's own favor because the pseudo-Socratic statement "Nobody admits wrong

²¹⁶ *Op. cit.*, p. 301.

²¹⁷ Of course, a person may *both* be making an exception in his own favor *and also* be challenging a moral rule; but the latter is not a serious challenge.

²¹⁸ "Impartiality and Consistency" (1961).

willingly" happens to be true—we are all "in" the moral game, and for this reason can be convicted of inconsistency of the will.

§10. FREEDOM AND THE HIGHEST GOOD

Of all the many aspects of Kant's philosophy, his treatment of freedom shows the longest development period, and many of his final conclusions were not reached until the *Religion* or the *Metaphysic of Morals*: later development has still not been adequately studied. This becomes apparent in Schrader's treatment of a question about the second *Critique* that he deals with in what is less a review than an important investigation of two questions prompted by Beck's commentary on the *Critique of Practical Reason*. The first of these questions is why the Aesthetic plays such a subordinate role in the ethical work, and for this Schrader suggests three reasons:

Kant regarded practical reason as determining the will directly and in this respect radically different from theoretical reason, which is related to objects only through the mediation of the senses. But a second reason is that, according to Kant, practical reason produces its object. Since practical reason is not concerned with what is . . . but with what ought to be, it is not dependent upon anything comparable to sensible intuition, and hence requires no aesthetic foundation.²¹⁹

Finally Kant was more concerned in the second *Critique* to establish the purity of pure reason than to limit its claims to autonomy, and so it is not concerned to show that "practical reason is essential for and constitutive of empirical desire as such. The argument of the Analytic of the *Critique of Practical Reason* thus differs basically from that of the first *Critique*."²²⁰

These three reasons account for a certain lack of balance in the second *Critique*, and others, arising from Kant's historical situation, could be suggested. This lack of balance Schrader finds in the absence of any doctrine analogous to the necessary unity of apperception, for:

if the "I think" must accompany every sensation, the "I will" or its equivalent must surely accompany

every empirical desire. For a desire to be in any meaningful sense my desire, it must be integrally related to my consciousness.²²¹

Thus Schrader argues that we must not treat reason as "a separate non-empirical faculty which serves only to constrain and control empirical desire." For these and other reasons he concludes that an Aesthetic ought to be incorporated in the second *Critique*, which cannot rely on that of the first because "when the shift is made from the objective-cognitive perspective on the world to that of practical reason, there is a change of signature for all data to be considered."²²²

This lack of an Aesthetic is something which is, to some extent, made up in the later writings, where the will is more fully incorporated into the phenomenal world. To the difficulties of this incorporation Silber has drawn attention, arguing that Kant's attempt to resolve the problems of the Third Antimony through the distinction between phenomena and noumena is plainly inadequate, since it would demand what Kant cannot admit, namely, a pre-established harmony between the two worlds.²²³ An alternative solution is plainly necessary, he argues, because moral experience involves "the temporal awareness of duty and the involvement of the moral agent in the phenomenal world,"²²⁴ and suggests that such a solution might be discovered in the third *Critique*. We may agree that such a solution is desirable, although it is not altogether essential if, like many German commentators, we are prepared to take a tough metaphysical line. But that Kant did himself point the direction away from his theory of the first *Critique* is shown in the development of the concept of *Willkür* in the later ethical writings.

A good introduction to the reasons for which Kant introduced this concept is provided by Fackenheim's treatment of the *Religion*, which begins by pointing out how Kant, who had appealed to many of his contemporaries by showing how man could rise above nature, suddenly shocked them with his doctrine of the radical evil in human nature. Fackenheim traces the doctrine out of a deficiency in the views of the earlier ethical writings where Kant maintains that "the

²¹⁹ "Basic Problems of Philosophical Ethics" (1964), p. 105.

²²⁰ *Ibid.*, p. 106.

²²¹ *Ibid.*, p. 110.

²²² *Ibid.*, p. 109.

²²³ "The Ethical Significance of Kant's 'Religion'" (1960), p. xcvi.

²²⁴ *Ibid.*, p. 101; cf. Beck, *Commentary* (1960), p. 188.

will can determine itself in one way only: towards obedience to the moral law . . . Hence a free will can only be a good will."²²⁵ In this he was moving in the van of the Platonic-Plotinian tradition embodied, for example, in Whichcote, who maintained that "he is least of all free; nay, he is the veriest slave in the world; who hath either will or power to vary from the law of right."²²⁶ But this view "is compelled to deny that there can be such a thing as an evil will. Along with the evil will it must deny evil itself. And in denying both, it cannot justify moral responsibility for moral evil."²²⁷ After the *Grundlegung*, however, Kant struggles toward the conception of heteronomy as a mode of willing, which will make the will definable in terms of both desire and practical reason; however, this has:

the dismayingly consequence that a person is still a person in possession of his freedom even if he rejects the law. Thus the law no longer appears to be related to the will as a condition of its being. The categorical imperative seems to resolve itself into a hypothetical one: if one wishes to be moral he must obey the moral law. . . .²²⁸

This consequence leads to the equivocation of the second *Critique*, and is only settled with the introduction of the distinction between *Wille* and *Willkür*, and the concept of *Gesinnung*.

The detailed articulation of these concepts is still a matter of debate. Fackenheim does little to clarify them, in that it is not clear in his account how Kant gets his disjunction that man must be either radically evil or radically good out of his statements about the relation of motives and dispositions, nor is it clear how Fackenheim thinks Kant reconciles the statements that radical evil is both innate and is also brought upon man by himself. There is, nevertheless, a clear distinction drawn between Kant's doctrine of the radical evil in human nature and the Christian doctrine of original sin: for Kant there is the possibility of "a kind of rebirth, as it were a new creation."²²⁹ There is the possibility of such recreation in every act of

decision, but this will be "an ultimate act of decision for which there is no higher ground."²³⁰ Man is no longer free only in so far as he is good: he is also free to choose evil.

Nevertheless, the older concept of freedom still finds its place in the later theories. It could be said that man is free to choose to be free. In its first occurrence, "free" refers to the activity of *Willkür*, which is able to choose to follow the dictates of inclination or of reason. In its second occurrence, "free" refers to the consideration that man is only really free when *Willkür* is determined by *Wille*. This last phrase is, Silber argues, a misleading one, in that, strictly, the *Wille* does not act at all.²³¹ It provides an incentive which, if strong enough, is "adopted by *Willkür* into the maxim of its choice," in which case we may speak of the *Willkür* as autonomous. Both *Wille* and *Willkür* are spontaneous in their activity, although in different ways, but it is only in *Willkür's* submission to, or free election of, the dictate of *Wille* that man may be spoken of as behaving autonomously. The matter is excellently treated by both Silber and Beck, and although there are disagreements between them,²³² these seem to be largely because Beck occasionally uses the word "free" equivocally, leaving the sense involved to be gathered from the context.

As interesting as Kant's distinction of *Wille* and *Willkür* is Kant's account of their relationship in terms of the ability of *Willkür* to elect not only individual acts but also dispositions. Philosophically the relation of disposition to individual performance is one of extreme complexity, and it is therefore scarcely surprising that Kant, more in the interests of simplicity than of truth, speaks of there being a single disposition of which the sum of any person's acts are, in some way, an expression. The *Willkür* can follow respect for the dictates of *Wille* into both making the disposition to goodness its primary maxim, and also in following this disposition through in a series of acts which, through their conformity with the law, will reflect the dispositional act. Alternatively, *Willkür* can descend

²²⁵ "Kant and Radical Evil" (1954), p. 344; cf. Silber, *op. cit.*, pp. lxxxii-ii.

²²⁶ *Aphorisms* (1753) A725; cf. *Sermons* (1698), p. 307.

²²⁷ *Op. cit.*, p. 345.

²²⁸ Silber, *op. cit.*, p. lxxxv.

²²⁹ Silber, *Religion Within the Limits of Reason Alone*, trans. Greene and Hudson (1960), p. 43. This new edition does the great service of marking Kant's distinction of *Wille* and *Willkür*, but it seems to me regrettable that opportunity was not taken for incorporating the Prussian Academy pagination.

²³⁰ Fackenheim, *op. cit.*, p. 352.

²³¹ *Op. cit.*, pp. civ-cv; cf. Beck, *Commentary* (1960), p. 180.

²³² Cf. Silber, "The Importance of the Highest Good in Kant's Ethics" (1962), p. 181.

into sin through frailty, which is a conscience stricken violation of the dispositional act, or through impurity of the will, where respect for the law is not the all sufficient motive of action, and conscience is frequently silent, or, finally, through the dispositional choice of evil, where the law is constantly subordinated to inclination. But if, *per impossibile*, the voice of conscience is entirely stilled, "the free *Willkür* loses its freedom altogether, and becomes mere animal '*Willkür*,'" and so, in the concepts of freedom and of personality, we obtain the necessary connection between the will and the categorical imperative which shows how the categorical imperative is possible.²³³ We must be grateful to Silber for making us aware of the importance and difficulty of the *Religion*, which is virtually the only place that Kant discussed the continuous aspects of personality and will.

Silber has also been publishing several studies of the concept of the highest good in Kant's ethics that are intrinsically related to his interest in the *Religion*. He has pointed out that Kant was inconsistent in maintaining the following three statements:

- (1) Man is morally obligated to attain in full the highest good.
- (2) That to which man is obliged must be possible.
- (3) The full attainment of the highest good by man is, in fact, not possible.²³⁴

Silber argues that it was the inconsistency of these three that led Kant to postulate God and immortality as necessary conditions of the moral law, but he also argues that even with these postulates they are not rendered consistent. If we concede this, we must also concede that one of (1)-(3) is false. That (3) might be false is rejected out of hand, and that (2) is false is argued against on the grounds that "ought" implies "can" only because "can" is a presupposition of "ought." Silber's choice thus falls on (1), which must be changed to say only that man is obligated "to approximate the highest good to the fullest possible degree."²³⁵ In this way the importance becomes

apparent of Kant's statement that "nothing is more reprehensible than to derive the laws prescribing what ought to be done from what is done, or to impose upon them the limits by which the latter is circumscribed":²³⁶ the highest good becomes a regulative idea against which we can estimate our freedom.

Essentially the same ground is covered in the second question which Schrader asks in his review of Beck. Agreeing with Beck that there is no proper Dialectic of practical reason, he asks: Is there a moral dialectic as such? If we were to argue by analogy from the first *Critique*, which is concerned with the finitude of human understanding, we should expect some similar tensions involving the finitude of the human will. Schrader argues in the direction of Kierkegaard that the categorical imperative is not "the injunction to will the law but rather to will in conformity with the demands of the law and hence to will an objective state of affairs."²³⁷ But the highest good, as the final object of man's moral will, is impossible of attainment, and man's will tends to fall to moral futility in the face of his impossibility.

At times [Kant] was inclined to accept a stoical resolution of the problem which would put the stress upon moral volition as a second-intensional act directed toward the will which is engaged in world process. But at other times he struggled with the problem of the realisability of the moral good as an objective state of affairs beyond the power of human freedom to achieve.²³⁸

Hence, he argues, God is needed within the Kantian system not only to reconcile virtue and happiness, but also "to assure that a moral order could be achieved through the exercise of human freedom."²³⁹

The account is valuable and interesting, while two related objections can be made to it. First, that it is impossible, even with God's aid, that a moral order could be realized through one man's will, and God may only be introduced in this way to support the incentive of the moral law, so that man may be engaged in the world to the limits of his ability. Second, the antithesis of "making pos-

²³³ Silber, "The Ethical Significance of Kant's Religion" (1960), p. cxxiv.

²³⁴ "Kant's Conception of the Highest Good as Immanent and Transcendent" (1959), p. 479.

²³⁵ *Ibid.*, p. 478.

²³⁶ A919=B375.

²³⁷ "Basic Problems of Philosophical Ethics" (1964), p. 113.

²³⁸ *Ibid.*, p. 115.

²³⁹ *Ibid.*, p. 116.

sible a reconciliation of virtue and happiness" and "making possible a moral order" is one to which some small exception may be made.

This may be seen through two of Silber's articles on the highest good. In the first of these²⁴⁰ Silber is concerned with the statement that "good and evil must be defined after and by means of the law."²⁴¹ He is therefore largely concerned with Theorems I-III of the second *Critique*, and his expert exposition culminates in the statement:

The attempt to ground the principle of morality on a previously defined material concept of good founders on this dilemma: either the good stands in no relation or in a contingent relation to the will, or the good itself has the power to determine the will to action and thereby destroys itself. In neither case can the moral law be derived from the good and, therefore, no relation of obligation can be effected between the good and the will.²⁴²

The dilemma is as sound as Silber's exegesis, which should be compared with Beck's account, that supplements it at many points,²⁴³ but which has also to be brought into relation with recent attacks on Theorem II, e.g. by Reiner.²⁴⁴ The base of his argument is also, however, extended by Silber to include facts of our moral experience which are incompatible with many moral theories. Thus, if good is defined prior to the law "it becomes a homogeneous concept and is related to the will as the object of its desire. But if the good is the object of desire the good is always sought, and virtue and happiness become identified since to the extent that one attains the good he is both virtuous and happy."²⁴⁵ Emphasis must therefore be placed upon the heterogeneity of the good: upon its unification of the formal and the material which, elsewhere, Silber argues as essential to the Kantian theory.²⁴⁶ In a third article Silber specifies these formal and material elements more precisely. As one element of the highest good we have the good will, which "is itself the object of the will, and in its act of volition it wills nothing more or less than its

own perfection (free willing) as an end which is also a duty."²⁴⁷ Contingent upon this there is the material demand, which is the foundation for duties to oneself, for the "functional completeness of all our powers" which are necessary for, in so far as they are necessary for, the exercise of moral volition. Yet further content needs to be given if we are to determine the concept of the highest good, and Kant therefore insists that "one is likewise obligated to seek the happiness of others as a second end which is also a duty."²⁴⁸ The highest good is produced through the combination of these two ends: it is the:

synthesis of the moral good and the natural good. And since the moral good is the supreme condition of this unity, we find that in the fulfilment of the highest good happiness must be present in exact proportion to morality.²⁴⁹

It is the balance of happiness and worthiness to be happy that distinguishes Kant's good from many utilitarian ideals, and it is something which we can go some way toward achieving, although Kant argues that God, and an after life, are necessary for the final balancing of the books.²⁵⁰ It is interesting to notice that the appeal to God is made not only because we are unable to balance the books, but also that a part reason for this is that we have no sure method of checking the balance sheets. We have the two propositions (1) that happiness ought to be distributed according to virtue, and (2) that virtue is not to be assessed in terms of its effects. The problem has its modern legacy in a distinction which we may make between moral optimists and moral pessimists. The first are those who believe that actions are the best index of virtue that we possess, and therefore advocate distributing the material component of happiness accordingly. The moral pessimists, on the other hand, are those who argue for equal distribution on the grounds either (a) that our safest course is to assume virtue to be distributed equally, or (2) that although virtue is not equally distributed it is, in any case, its own

²⁴⁰ "The Copernican Revolution in Ethics: the Good Re-examined" (1959-60).

²⁴¹ Ak. V. 63.

²⁴² *Op. cit.*, pp. 86-87.

²⁴³ *Commentary* (1960), pp. 95ff.

²⁴⁴ "Kants Beweis zur Widerlegung des Eudämonismus und das A Priori der Sittlichkeit" (1962-3).

²⁴⁵ *Op. cit.*, p. 98.

²⁴⁶ "The Context of Kant's Ethical Thought" (1959), Pt. II.

²⁴⁷ "The Importance of the Highest Good in Kant's Ethics" (1962), p. 186.

²⁴⁸ *Ibid.*, p. 191.

²⁴⁹ *Ibid.*, p. 193.

²⁵⁰ Cf. Silber, "The Metaphysical Importance of the Highest Good etc." (1959).

reward, or, less consequentially although primarily, (3) that there is no such thing as freedom or virtue. Although I have drawn them crudely, these are

opposed positions which are the modern inheritance of the problem that Kant attempted to solve with his moral argument for the existence of God.

University of York

SELECTED BIBLIOGRAPHY

- ADDIS, L. C. "Kant's First Analogy," *Kant-Studien*, vol. 54 (1963), pp. 237-242.
- ALBRECHT, W. "Die sogenannte neue Deduktion in Kants Opus Postumum," *Archiv für Philosophie*, vol. 8, (1954), pp. 57-65.
- ALEXANDER, H. G. "Necessary Truth," *Mind*, vol. 66 (1957), pp. 507-521.
- ALEXANDRE, Michel *Lecture de Kant* (Paris, 1961).
- ANCESCHI, Luciano *D. Hume e i presupposti empiristici della estetica Kantiana* (Milano, 1956).
- ANTONOPOULOS, Georg *Der Mensch als Bürger zweier Welten. Ein Beitrag zur Entwicklungsgeschichte von Kants Philosophie* (Bonn, 1958).
- AXINN, Sidney "Kant, Logic, and the Concept of Mankind," *Ethics*, vol. 48 (1958), pp. 286-291.
- "And Yet: A Kantian Analysis of Aesthetic Interest," *Philosophy and Phenomenological Research*, vol. 25 (1964), pp. 108-116.
- BAIER, Kurt *The Moral Point of View* (Ithaca, Cornell, 1958).
- BALLARD, Edward G. "The Kantian Solution to the Problem of Man within Nature," *Tulane Studies in Philosophy*, vol. 3 (Hague, Nijhoff, 1954).
- BALLAUF, Theodor *Vernünftiger Wille und gläubige Liebe* (Meisenheim, Hain, 1957).
- BARBAR, Edward G. "Two Logics of Modality," *Tulane Studies in Philosophy*, vol. 3 (Hague, Nijhoff, 1954).
- BARONE, Francesco "Kant e la logica formale," *Filosofia*, vol. 7 (1956), pp. 697-750.
- "I problemi e il problema della logica trascendentale kantiana," *Filosofia*, vol. 8 (1957), pp. 19-68.
- BECK, Lewis White "Can Kant's Synthetic Judgments Be Made Analytic?" *Kant-Studien*, vol. 47 (1955-6), pp. 168-181.
- "Sir David Ross on Duty and Purpose in Kant," *Philosophy and Phenomenological Research*, vol. 41 (1955), pp. 98-107.
- "Kant's Theory of Definition," *The Philosophical Review*, vol. 65 (1956), pp. 179-191.
- "Apodictic Imperatives," *Kant-Studien*, vol. 49 (1957), pp. 7-24.
- "The Meta-semantics of the Problem of the Synthetic *a Priori*," *Mind*, vol. 66 (1957), pp. 228-232.
- *Commentary on Kant's Critique of Practical Reason* (Chicago, 1960).
- "Das Faktum der Vernunft. Zur Rechtfertigungsproblematik in der Ethik," *Kant-Studien*, vol. 52 (1960-1), pp. 271-282.
- BENNETT, Jonathan "The Analytic and the Synthetic," *Proceedings of the Aristotelian Society*, vol. 59 (1958-9), pp. 163-188.
- "The Status of Determinism," *The British Journal for the Philosophy of Science*, vol. 14 (1963), pp. 106-119.
- BERNAYS, Paul "Zur Frage der Anknüpfung an die Kantische Erkenntnistheorie," *Dialectica*, vol. 9 (1955), pp. 23-65; 195-221.
- BETH, E. W. "Kants Einteilung der Urteile in analytische und synthetische," *Algemeen Nederlands Tijdschrift voor Wijsbegeerte en Psychologie*, vol. 46 (1954), pp. 253-264.
- BIEMEL, W. *Die Bedeutung von Kants Begründung der Ästhetik für die Philosophie der Kunst* (*Kant-Studien Ergänzungshefte*, Nr. 77, 1959).
- BIRD, Graham "The Necessity of Kant," *Mind*, vol. 68 (1959), pp. 389-392.
- "Analytic and Synthetic," *The Philosophical Quarterly*, vol. 11 (1961), pp. 227-237.
- *Kant's Theory of Knowledge* (London, Routledge & Kegan Paul, 1962).
- BLACHOS, Georges K. *La Pensée politique de Kant* (Paris, 1962).
- BROWN, Stuart M. "Has Kant a Philosophy of Law?" *The Philosophical Review*, vol. 71 (1962), pp. 33-48.
- BUTTS, Robert E. "Hypothesis and Explanation in Kant's Philosophy of Science," *Archiv für Geschichte der Philosophie*, vol. 43 (1961), pp. 153-170.
- "Kant on Hypotheses in the 'Doctrine of Method' and the *Logik*," *Archiv für Geschichte der Philosophie*, vol. 44 (1962), pp. 185-203.
- CAMPO, M. *Schizzo storica della esegesi e critica kantiana dal "ritorno a Kant" alla fine dell'Ottocento* (Varese, 1959).
- CASULA, M. *Marechal e Kant* (Roma, 1955).
- *Studi kantiani sul trascendente* (Milano, 1963).
- CHIODI, Pietro *La Deduzione nell'opera di Kant* (Torino, 1961).
- CONINCK, A. de *L'analytique transcendental de Kant* (Louvain, 1955).
- COUSIN, D. R. "Kant on the Self," *Kant-Studien*, vol. 49 (1957-8), pp. 25-35.
- CRAMER, W. *Die Monade. Das philosophische Problem vom Ursprung* (Stuttgart, 1954).

- CRAWFORD, Patricia A. "Kant's Theory of Philosophical Proof," *Kant-Studien*, vol. 53 (1961-2), pp. 257-268.
- DELEKAT, Friedrich *Immanuel Kant* (Heidelberg, Quelle & Meyer, 1963).
- DELIUS, Harald *Untersuchungen zur Problematik der sogenannten synthetischen Sätze a priori* (Göttingen, Vandenhoeck & Rupprecht, 1963).
- DERGGIBUS, A. *Il problema morale in J. J. Rousseau e la validità dell'interpretazione kantiana* (Torino, 1957).
- DIEMERS, A. "Zum Problem des Materialen in der Ethik Kants," *Kant-Studien*, vol. 45 (1953-4), pp. 21-32.
- DIETRICHSON, Paul "What Does Kant Mean by 'Acting from Duty'?" *Kant-Studien*, vol. 53 (1961-2), pp. 277-288.
- DUNCAN, A. R. C. *Practical Reason and Morality* (London, Nelson, 1957).
- EBBINGHAUS, Julius "Kants Ableitung des Verbotes der Lüge aus dem Rechte der Menschheit," *Revue internationale de Philosophie*, vol. 8 (1954), pp. 409-422.
- "Interpretation and Misinterpretation of the Categorical Imperative," (trans. H. J. Paton), *The Philosophical Quarterly*, vol. 4 (1954), pp. 97-108.
- ENGEL, S. Morris "Kant's Copernican Analogy: A Re-examination," *Kant-Studien*, vol. 54 (1962-3), pp. 243-251.
- "Kant's Refutation of the Ontological Argument," *Philosophy and Phenomenological Research*, vol. 24 (1963), pp. 20-35.
- "On the 'Composition' of the Critique: A Brief Comment," *Ratio*, vol. 6 (1964), pp. 81-91.
- EWING, A. C. "Kant's Attack on Metaphysics," *Revue internationale de Philosophie*, vol. 8 (1954), pp. 371-391.
- FACKENHEIM, Emil "Kant and Radical Evil," *University of Toronto Quarterly*, vol. 23 (1954), pp. 339-353.
- "Kant's Concept of History," *Kant-Studien*, vol. 48 (1956-7), pp. 381-398.
- FEIBLEMAN, James K. "Kant and Metaphysics," *Tulane Studies in Philosophy*, vol. 3 (Hague, Nijhoff, 1954).
- FRIEDMANN, Lawrence "Kant's Theory of Time," *The Review of Metaphysics*, vol. 7 (1954), pp. 379-388.
- FUNKE, Gerhard "Gewohnheit," *Archiv für Begriffsgeschichte*, vol. 3 (1958), pp. 479-496.
- GAHRINGER, E. "The Metaphysical Aspects of Kant's Moral Philosophy," *Ethics*, vol. 44 (1954), pp. 277-291.
- GRAYEFF, F. "The Relation of Transcendental and Formal Logic," *Kant-Studien*, vol. 51 (1959-60), pp. 349-352.
- GREGOR, Mary J. *Laws of Freedom* (Oxford, Blackwell, 1963).
- GRICE, H. P. & STRAWSON, P. F. "In Defence of a Dogma," *The Philosophical Review*, vol. 65 (1956), pp. 141-158.
- GOODMAN, Nelson "On Likeness of Meaning" *Analysis*, vol. 10 (1949), pp. 1-7.
- GUÉROULT, Martial "Canon de la raison pure et Critique de la raison pratique," *Revue internationale de Philosophie*, vol. 8 (1954), pp. 331-357.
- "Vom Kanon der Kritik der reinen Vernunft zur Kritik der praktischen Vernunft," *Kant-Studien*, vol. 54 (1963), pp. 432-444.
- HAEZRAHI, Pepita "The Avowed and Unavowed Sources of Kant's Theory of Ethics," *Ethics*, vol. 61 (1952), pp. 157-168.
- "The Concept of Man as an End in Himself," *Kant-Studien*, vol. 53 (1961-2), pp. 209-224.
- HALL, Robert W. "Kant and Ethical Formalism," *Kant-Studien*, vol. 52 (1960-1), pp. 433-439.
- HAMBURG, Carl H. "Kant, Cassirer, and the Concept of Space," *Tulane Studies in Philosophy*, vol. 3 (Hague, Nijhoff, 1954).
- HAMLIN, D. W. "Analytic Truths," *Mind*, vol. 65 (1956), pp. 359-376.
- "On Necessary Truth," *Mind*, vol. 70 (1961), pp. 514-525.
- HANCOCK, Roger N. "Ethics and History in Kant and Mill," *Ethics*, vol. 68 (1957), pp. 56-60.
- "Kant and the Natural Right Theory," *Kant-Studien*, vol. 52 (1960-1), pp. 440-447.
- HANSON, N. R. "The Very Idea of a Synthetic a priori," *Mind*, vol. 71 (1962), pp. 521-524.
- HARRISON, Jonathan "Kant's Examples of the First Formulation of the Categorical Imperative," *The Philosophical Quarterly*, vol. 7 (1957), pp. 50-62.
- "The Categorical Imperative," *The Philosophical Quarterly*, vol. 8 (1958), pp. 360-364.
- HEIDEMANN, Ingeborg "Der Begriff der Spontaneität in der Kritik der reinen Vernunft," *Kant-Studien*, vol. 47 (1955-6), pp. 3-30.
- "Person und Welt. Zur Kantinterpretation von Heinz Heimsoeth," *Kant-Studien*, vol. 48 (1956-7), pp. 344-360.
- "Zur Kantforschung von H. J. Paton," *Kant-Studien*, vol. 49 (1957-8), pp. 107-142.
- *Spontaneität und Zeitlichkeit* (*Kant-Studien Ergänzungshefte*, Nr. 75, 1958).
- HEIMSOETH, Heinz *Studien zur Philosophie Immanuel Kants* (*Kant-Studien Ergänzungshefte*, Nr. 71, 1956). Contains, with dates of first publication:
- (1) "Chr. Wolffs Ontologie und die Prinzipienforschung I. Kants" (1956).
 - (2) "Der Kampf um den Raum in der Metaphysik der Neuzeit" (1925).
 - (3) "Metaphysik und Kritik bei Chr. A. Crusius" (1926).
 - (4) "Metaphysische Motive in der Ausbildung des kritischen Idealismus" (1925).

- (5) "Persönlichkeitsbewusstsein und Ding an sich in der kantischen Philosophie" (1924).
- "Vernunftantinomie und transzendente Dialektik in der geschichtlichen Situation des kantischen Lebenswerkes," *Kant-Studien*, vol. 51 (1959-60), pp. 131-141.
- *Atom, Seele, Monade* (Mainz Akademie der Wissenschaften und der Literatur; Abhandlungen der Geistes und Sozialwissenschaften Klasse, Nr. 3, 1960).
- *Studien zur Philosophiegeschichte* (*Kant-Studien Ergänzungshefte*, Nr. 82, 1961). Contains, among much else:
- (1) "Zur Geschichte der Kategorienlehre" (1952).
 - (2) "Zeitliche Weltunendlichkeit und das Problem des Anfangs" (1960).
 - (3) "A. Colliers 'Universal Schlüssel' und der Durchbruch des neuzeitlichen Bewusstseinsidealismus" (1960).
- "Zur Herkunft und Entwicklung von Kants Kategorientafel," *Kant-Studien*, vol. 54 (1962-3), pp. 376-403.
- *Astronomisches und Theologisches in Kants Weltverständnis* (Mainz Akademie der Wissenschaften und der Literatur; Abhandlungen der Geistes und Sozialwissenschaften Klasse, Nr. 9, 1963).
- HENDEL, C. W. (Ed.) *The Philosophy of Kant and Our Modern World* (New York, Liberal Arts, 1957).
- HENRICH, Dieter "Hutcheson and Kant," *Kant-Studien*, vol. 49 (1957-8), pp. 49-69.
- "Der Begriff der sittlichen Einsicht und Kants Lehre vom Faktum der Vernunft," in *Die Gegenwart der Griechen in neuen Denken* (Tübingen, Mohr, 1960).
- "Über Kants früheste Ethik," *Kant-Studien*, vol. 54 (1963), pp. 404-431.
- *Der ontologische Gottesbeweis* (Tübingen, Mohr, 1960).
- HERRING, Herbert *Das Problem der Affektion bei Kant* (*Kant-Studien, Ergänzungshefte*, Nr. 67, 1953).
- "Leibniz' principium identitatis indiscernibilium und die Leibniz-Kritik Kants," *Kant-Studien*, vol. 49 (1957-8), pp. 389-400.
- HILDEBRANDT, Kurt "Kants Verhältnis zu Leibniz in der vorkritischen Periode," *Zeitschrift für philosophische Forschung*, vol. 8 (1954), pp. 3-29.
- HÜBNER, K. "Leib und Erfahrung in Kants Opus Postumum," *Zeitschrift für philosophische Forschung*, vol. 7 (1953), pp. 204-219.
- JANOSKA, Georg "Der transzendente Gegenstand," *Kant-Studien*, vol. 46 (1954-5), pp. 193-221.
- KAHL-FURTHMANN, G. "Subjekt und Objekt; Ein Beitrag zur Vorgeschichte der Kant'schen Kopernikanischen Wendung," *Zeitschrift für philosophische Forschung*, vol. 7 (1953), pp. 326-339.
- KAMINSKY, J. "Kant's Analysis of Aesthetics," *Kant-Studien*, vol. 50 (1958-9), pp. 77-88.
- KAULBACH, Friedrich "Kants Beweis des 'Daseins der Gegenstände im Raum ausser mir'," *Kant-Studien*, vol. 50 (1958-9), pp. 323-347.
- "Geist und Raum," *Wissenschaft und Weltbild*, vol. 12 (1959), pp. 523-533.
- *Die Metaphysik des Raumes bei Leibniz und Kant* (*Kant-Studien Ergänzungshefte*, Nr. 79, 1960).
- "Das Prinzip der Bewegung in der Philosophie Kants," *Kant-Studien*, vol. 54, (1963), pp. 3-16.
- "Leibbewusstsein und Welterfahrung beim frühen und späten Kant," *Kant-Studien*, vol. 54 (1963), pp. 464-490.
- *Der Philosophische Begriff der Bewegung* (Köln, Böhlau, 1965).
- KEMP, J. "Kant's Examples of the Categorical Imperative," *The Philosophical Quarterly*, vol. 8 (1958), pp. 63-71.
- *Reason, Action, and Morality* (London, Routledge & Kegan Paul, 1964).
- KLAUSEN, Sverre *Kants Ethik und ihre Kritiker* (Oslo, 1954).
- *Grundgedanken der materialen Wertethik bei Hartmann und Scheler in ihrem Verhältnis zur Kantischen* (Oslo, 1958).
- *Das Problem der Erkennbarkeit der Existenz Gottes bei Kant* (Oslo, 1959).
- KNITTERMEYER, H. "Zu Heinz Heimsoeths Kantdeutung," *Kant-Studien*, vol. 49 (1957-8), pp. 293-311.
- KNOX, T. M. "Hegel's Attitude to Kant's Ethics," *Kant-Studien*, vol. 49 (1957-8), pp. 70-81.
- KOLENDA, Konstantin "Professor Ebbinghaus' Interpretation of the Categorical Imperative," *The Philosophical Quarterly*, vol. 5 (1955), pp. 74-77.
- KONRAD, Johanna "Inwieweit hat Kants Personenbegriff Bedeutung und Gültigkeit für unsere Zeit?" *Jahrbuch der Albertus-Universität Königsberg/Pr.*, vol. 5 (1954), pp. 97-112.
- KOPPER, Joachim "Kants Gotteslehre," *Kant-Studien*, vol. 47 (1955-6), pp. 31-61.
- "Antwort an W. A. Schulze," *Kant-Studien*, vol. 48 (1956-7), pp. 84-85.
- KÖRNER, Stephan *Kant* (Penguin Books, 1955).
- *The Philosophy of Mathematics* (London, Hutchinson, 1960).
- LACHÈZE-REY, P. "Réflexions historiques et critiques sur la possibilité des jugements synthétiques à priori," *Revue internationale de Philosophie*, vol. 8 (1954), pp. 358-370.
- LANGE, H. "Über den Unterschied der Gegenden im Raume," *Kant-Studien*, vol. 50 (1958-9), pp. 479-499.
- LEE, Harold N. "The Rigidity of Kant's Categories," *Tulane Studies in Philosophy*, vol. 3 (Hague, Nijhoff, 1954).
- LEHMANN, Gerhard "Erscheinungsstufung und Realitätsproblem in Kants Opus Postumum," *Kant-Studien*, vol. 45 (1953-4), pp. 140-154.
- "Kritizismus und kritische Motive in der Entwicklung der kantischen Philosophie," *Kant-Studien*, vol. 48 (1956-7), pp. 25-54.

- "Voraussetzungen und Grenzen systematischer Kantinterpretation," *Kant-Studien*, vol. 49 (1957-8), pp. 364-388.
- "System und Geschichte in Kants Philosophie," *Il Pensiero*, vol. 3 (1958).
- "Kants Widerlegung des Idealismus," *Kant-Studien*, vol. 50 (1958-9), pp. 348-362.
- "Zur Problemanalyse von Kants Nachlasswerk," *Il Pensiero*, vol. 6 (1961).
- "Zur Frage der Späntwicklung Kants," *Kant-Studien*, vol. 54 (1963), pp. 491-507.
- LOTZ, J. B. "Die Raum-Zeit Problematik in Auseinandersetzung mit Kants transzendentaler Ästhetik," *Zeitschrift für philosophische Forschung*, vol. 8 (1954), pp. 30-43.
- (Ed) *Kant und die Scholastik Heute* (Pullach, Berchmanskolleg, 1955).
- LUPORINI, Cesare *Spazio e materia in Kant* (Firenze, 1961).
- MALGAUD, W. "Kants Begriff der empirischen Realität," *Kant-Studien*, vol. 54 (1963), pp. 288-303.
- MARC-WOGAU, Konrad "Kants Lehre vom analytischen Urteil," *Theoria*, vol. 42 (1951), pp. 140-154.
- MARQUARDT, C. *Skeptische Methode in Blick auf Kant* (Freiburg, 1958).
- MARTIN, Gottfried *Kant's Metaphysics and Theory of Science* (tr. P. G. Lucas, Manchester, 1955).
- *Gesammelte Abhandlungen und Vorträge I (Kant-Studien Ergänzungshefte, Nr. 81, 1961)*.
- "Probleme der Prinzipienlehre in der Philosophie Kants," *Kant-Studien*, vol. 52 (1960-1), pp. 173-184.
- MATHIEU, Vittorio "La deduzione trascendentale di Kant," *Filosofia*, vol. 7 (1956), pp. 405-440.
- *La filosofia trascendentale e l'Opus Postumum di Kant* (Torino, 1958).
- MATSON, W. L. "Kant as Casuist," *The Journal of Philosophy*, vol. 51 (1954), pp. 855-890.
- MAYO, Bernard "Incongruity of Counterparts," *Philosophy of Science*, vol. 25 (1958), pp. 109-115.
- McRAE, Robert "Kant's Conception of the Unity of the Sciences," *Philosophy and Phenomenological Research*, vol. 18 (1957), pp. 1-17.
- MILMED, Bella K. *Kant and Current Philosophical Issues* (New York, 1961).
- MONRO, D. H. "Impartiality and Consistency," *Philosophy*, vol. 36 (1961), pp. 161-176.
- MORITZ, Manfred *Kants Einteilung der Imperative* (Copenhagen, Munksgaard, 1960).
- MÜLLER-LAUTER, Wolfgang "Kants Widerlegung des materialen Idealismus," *Archiv für Geschichte der Philosophie*, vol. 46 (1964), pp. 60-82.
- MURALT, A. de *La conscience transcendente dans le criticisme kantien* (Paris, 1958).
- NAHM, Milton C. "Sublimity and the 'Moral Law' in Kant's Philosophy," *Kant-Studien*, vol. 48 (1957-8), pp. 502-524.
- NAWRATIL, K. "Wie ist Metaphysik nach Kant möglich?" *Kant-Studien*, vol. 50 (1958-9), pp. 163-177.
- NEGRI, Antonio *La comunità estetica in Kant* (Galatina, 1957).
- *Alle origini del formalismo giuridico. Studio sul problema della forma in Kant e nei giuristi kantiani tra il 1789 e il 1802* (Padova, 1962).
- PANETH, F. A. "Die Erkenntnis des Weltbaus durch Thomas Wright und Immanuel Kant," *Kant-Studien*, vol. 47 (1955-6), pp. 337-349.
- PARKINSON, G. H. R. "Necessary Propositions and A Priori Knowledge in Kant," *Mind*, vol. 69 (1960), pp. 391-397.
- PARSONS, Charles "Infinity and Kant's Conception of a Possible Experience," *The Philosophical Review*, vol. 73 (1964), pp. 182-197.
- PASINI, D. *Diritto, società e stato in Kant* (Milano, 1957).
- PATON, Herbert James *Kant's Metaphysics of Experience* (London, Allen & Unwin, 1936).
- "An Alleged Right to Lie: A Problem in Kantian Ethics," *Kant-Studien*, vol. 45 (1953-4), pp. 190-203.
- "Kant on Friendship," *Proceedings of the British Academy*, vol. 42 (1956), pp. 45-66.
- "The Aim and Structure of Kant's *Grundlegung*," (review of Duncan, *op. cit.*), *The Philosophical Quarterly*, vol. 8 (1958), pp. 112-130.
- "Formal and Transcendental Logic," *Kant-Studien*, vol. 49 (1957-8), pp. 245-263.
- PEACH, B. "Common Sense and Practical Reason in Reid and Kant," *Sophia*, vol. 24 (1956), pp. 66-71.
- PEARS, D. "The Incongruity of Counterparts," *Mind*, vol. 61 (1952), pp. 78-81.
- PELLEGRINO, U. *L'Ultimo Kant. Saggio critica sull'Opus Posthumum* (Milano, 1957).
- PLAASS, Peter *Kants Theorie der Naturwissenschaft* (Göttingen, Vandenhoeck & Rupprecht, 1965).
- QUINE, W. V. O. "Two Dogmas of Empiricism," *The Philosophical Review*, vol. 60 (1951), pp. 20-43; reprinted in *From a Logical Point of View* (Harvard, 1953), pp. 20-46.
- QUINTON, Anthony "Spaces and Times," *Philosophy*, vol. 37 (1962), pp. 130-174.
- REDMAN, Horst G. *Gott und Welt, Die Schöpfungstheologie der vorkritischen Periode Kants* (Göttingen, 1962).
- REINER, Hans *Pflicht und Neigung* (Meisenheim, Hain, 1951).
- "Die Goldene Regel," *Zeitschrift für philosophische Forschung*, vol. 3 (1948), pp. 74-105.
- "Kants Beweis zur Widerlegung des Eudämonismus und das A Priori der Sittlichkeit," *Kant-Studien*, vol. 54 (1962-3), pp. 129-165.
- RESCHER, Nicholas "Presuppositions of Knowledge," *Revue internationale de Philosophie*, vol. 13 (1959), pp. 418-429.

- RICHMAN, R. J. "Why Are Kant's Synthetic *A Priori* Judgments Necessary?" *Theoria*, vol. 30 (1964), pp. 5-20.
- ROBINSON, Richard "Necessary Propositions," *Mind* vol. 67 (1958), pp. 289-304.
- ROSS, David *Kant's Ethical Theory* (Oxford, Clarendon, 1954).
- ROTENSTREICH, Nathan "Kants Dialectic," *The Review of Metaphysics*, vol. 7 (1954), pp. 389-421.
- "Kant's Concept of Metaphysics," *Revue internationale de Philosophie*, vol. 8 (1954), pp. 392-408.
- "Kant's Schematism in Its Context," *Dialectica*, vol. 10 (1956), pp. 9-30.
- *Experience and its Systematization* (Hague, Nijhoff, 1965).
- RUST, Hans "Kritisches zu Kants Religionskritik," *Jahrbuch der Albertus-Universität zu Königsberg/Pr*, vol. 6 (1955), pp. 73-106.
- SANTELER, J. *Die Grundlegung der Menschenwürde bei I. Kant* (Innsbruck, 1962).
- SCHAPER, Eva "Kant's Schematism Reconsidered," *The Review of Metaphysics*, vol. 18 (1964), pp. 267-292.
- "The Kantian 'As-if' and Its Relevance for Aesthetics," *Proceedings of the Aristotelian Society*, vol. 65 (1965), pp. 219-234.
- SCHIPPER, E. W. "Kant's Answer to Hume's Problem," *Kant-Studien*, vol. 53 (1961-2), pp. 68-74.
- SCHMUCKER, Josef "Der Einfluss des Newtonschen Weltbildes auf die Philosophie Kants," *Philosophische Jahrbuch*, vol. 65 (1951), pp. 52-59.
- "Der Formalismus und die materialen Zweckprinzipien in der Ethik Kants," *Kant und die Scholastik Heute* (Ed. Lotz, q.v.), pp. 155-205.
- *Die Ursprünge der Ethik Kants* (Meisenheim, Hain, 1961).
- "Die Gottesbeweise beim vorkritischen Kant," *Kant-Studien*, vol. 54 (1963), pp. 445-463.
- SCHNEEBERGER, Guido *Kants Konzeption der Modalbegriffe* (Basel, 1952).
- SCHOELER, W. F. *Die transzendente Einheit der Apperzeption von Immanuel Kant* (Bern, 1959).
- SCHOLZ, Heinrich "Eine Topologie der Zeit im Kantischen Sinne," *Dialectica*, vol. 9 (1955), pp. 66-113.
- SCHRADER, George "Kant's Presumed Repudiation of the 'Moral Argument' in the Opus Postumum: An Examination of Adickes' Interpretation," *Philosophy*, vol. 26 (1951), pp. 228-241.
- "The Transcendental Ideality and Empirical Reality of Kant's Space and Time," *The Review of Metaphysics*, vol. 4 (1951), pp. 507-536.
- "The Status of Teleological Judgement in the Critical Philosophy," *Kant-Studien*, vol. 45 (1953-4), pp. 204-235.
- "Kant's Theory of Concepts," *Kant-Studien*, vol. 49 (1957-8), pp. 264-278.
- "Ontology and the Categories of Existence," *Kant-Studien*, vol. 54 (1963), pp. 47-62.
- "Autonomy, Heteronomy, and Moral Imperatives," *The Journal of Philosophy*, vol. 60 (1963), pp. 65-77.
- "Basic Problems of Philosophical Ethics," *Archiv für Geschichte der Philosophie*, vol. 46 (1964), pp. 102-117.
- SCHULZE, W. A. "Zu Kants Gotteslehre," *Kant-Studien*, vol. 48 (1956-7), pp. 80-84.
- SCHWARZ, Wolfgang "Kant's Philosophy of Law and International Peace," *Philosophy and Phenomenological Research*, vol. 23 (1962), pp. 71-80.
- SILBER, J. R. "The Metaphysical Importance of the Highest Good as the Canon of Pure Reason in Kant's Philosophy," *Texas Studies in Literature and Language*, vol. 1 (1959), pp. 233-244.
- "Kant's Conception of the Highest Good as Immanent and Transcendent," *The Philosophical Review*, vol. 68 (1959), pp. 469-492.
- "The Copernican Revolution in Ethics: the Good Re-examined," *Kant-Studien*, vol. 51 (1959-60), pp. 85-101.
- "The Context of Kant's Ethical Thought," *The Philosophical Quarterly*, vol. 9 (1959), pp. 193-207, 309-318.
- "The Ethical Significance of Kant's Religion," *Religion within the Limits of Reason Alone* (trans. Greene and Hudson, New York, Harper, 1960), pp. lxxix-cxxxiv.
- "Die Analyse des Pflicht- und Schuld- Erlebnisses bei Kant und Freud," *Kant-Studien*, vol. 52 (1960-1), pp. 295-309.
- "The Importance of the Highest Good in Kant's Ethics," *Ethics*, vol. 73 (1962), pp. 179-197.
- SINGER, Marcus G. "The Categorical Imperative," *The Philosophical Review*, vol. 63 (1954), pp. 577-591.
- *Generalisation in Ethics* (New York, Knopf, 1961).
- "The Golden Rule," *Philosophy*, vol. 38 (1963), pp. 293-314.
- SMART, Harold E. "Two Views on Kant and Formal Logic," *Philosophy and Phenomenological Research*, vol. 16 (1955), pp. 155-911.
- STEGMÜLLER, Wolfgang "Der Begriff des synthetischen Urteils *A Priori* und die moderne Logik," *Zeitschrift für philosophische Forschung*, vol. 8 (1954), pp. 535-563.
- STENIUS, Erik "On Kant's Distinction between Phenomena and Noumena," *Philosophical Essays Dedicated to Gunnar Aspelin* (Lund, 1963), pp. 230-246.
- "Are True Numerical Statements Analytic or Synthetic?" *The Philosophical Review*, vol. 74 (1965), pp. 357-372.
- STRAWSON, P. F. *Individuals* (London, Methuen, 1959).
- & GRICE, H. P. "In Defense of a Dogma," *The Philosophical Review*, vol. 65 (1956), pp. 141-158.

- SWINBURNE, R. G. "Times," *Analysis*, vol. 25 (1965), pp. 185-191.
- TONELLI, Giorgio "La formazione del testo della Kritik der Urteilskraft," *Revue internationale de Philosophie*, vol. 8 (1954), pp. 423-448.
- "Der Streit über die mathematische Methode in der Philosophie der ersten Hälfte des XVIII Jahrhunderts und die Entstehung von Kants Schrift über die 'Deutlichkeit'," *Archiv für Philosophie*, vol. 9 (1955), pp. 37-66.
- "Dall' estetica metafisica all' estetica psicoempirica. Studi sulle genesi del Criticismo (1754-1771) e sulle sue fonti (Memorie dell' Accademia delle Scienze di Torino, Ser. 3, Tom. 3, Pt. II, 1955).
- "L'origine della tavola dei giudici e del problema della deduzione delle categorie in Kant," *Filosofia*, vol. 7 (1956), pp. 129-138.
- "Von den verschiedenen Bedeutungen des Wortes 'Zweckmässigkeit' in der Kritik der Urteilskraft," *Kant-Studien*, vol. 49 (1957-8), pp. 154-166.
- "La tradizione delle categorie aristoteliche nella filosofia moderna sino a Kant," *Studi Urbinati*, Ser. B, vol. 32 (1958), pp. 121-143.
- *Elementi metafisici e metodologici in Kant (1754-1768)*, vol. 1 (Torino, 1959).
- "La question des bornes de l'entendement humain au XVIII^e siècle et la genèse du criticisme kantien," *Revue de Métaphysique et de Morale*, vol. 62 (1959), pp. 396-427.
- "Critiques of the Notion of Substance Prior to Kant," *Tijdschrift voor Philosophie*, vol. 23 (1961), pp. 285-301.
- "Der historische Ursprung der kantischen Termini 'Analytik' und 'Dialektik'," *Archiv für Begriffsgeschichte*, vol. 7 (1962), pp. 120-139.
- "Das Wiederaufleben der deutsch-aristotelischen Terminologie in der Entstehung der Kritik der reinen Vernunft," *Archiv für Begriffsgeschichte*, vol. 8 (1963), pp. 233-242.
- "Die Umwälzung von 1769 bei Kant," *Kant-Studien*, vol. 54 (1962-3), pp. 369-375.
- TURBAYNE, Colin M. "Kant's Refutation of Dogmatic Idealism," *The Philosophical Quarterly*, vol. 5 (1955), pp. 225-244.
- VALLENILLA, E. M. *El problema de la nada en Kant* (Madrid, 1965).
- VLEESCHAUWER, H. J. de *The Development of Kantian Thought* (trans. A. R. C. Duncan, London, Nelson, 1962).
- "A Survey of Kantian Philosophy," *The Review of Metaphysics*, vol. 11 (1957), pp. 122-142.
- "Etudes kantiennes contemporaines," *Kant-Studien*, vol. 54 (1962-3), pp. 63-119.
- "Wie ich jetzt die Kritik der reinen Vernunft Entwicklungsgeschichtlich lese," *Kant-Studien*, vol. 54 (1962-3), pp. 351-368.
- VLACHOS, Georges *La pensée politique de Kant* (Paris, 1963).
- VUILLEMIN, Jules *Physique et métaphysique kantiennes* (Paris, 1955).
- "Reflexionen über Kants Logik," *Kant-Studien*, vol. 52 (1960-1), pp. 310-335.
- WAGNER, Hans "Zur Kantinterpretation der Gegenwart, Rudolf Zocher und Heinz Heimsoeth," *Kant-Studien*, vol. 53 (1961-2), pp. 235-254.
- WAISMANN, F. "Analytic-Synthetic," *Analysis*, vol. 10 (1950), pp. 25-40.
- WALSH, W. H. "Schematism," *Kant-Studien*, vol. 49 (1957-8), pp. 95-106.
- "Kant's Moral Theology," *Proceedings of the British Academy*, vol. 49 (1963), pp. 263-289.
- WASSMER, Thomas A. "Responsibility and Pleasure in Kantian Morality," *Kant-Studien*, vol. 52 (1960-1), pp. 452-466.
- WEILER, Gershon "Kant's 'Indeterminate Concept' and the Concept of Man," *Revue internationale de Philosophie*, vol. 59 (1962), pp. 432-446.
- WEITZ, Morris "Analytic Statements," *Mind*, vol. 63 (1954), pp. 487-494.
- WENZL, Alois *Immanuel Kants bleibende Bedeutung* (München, 1954).
- WEYLAND, Klaus *Kants Geschichtsphilosophie (Kant-Studien, Ergänzungshefte, Nr. 85, 1963).*
- WHITE, Morton G. "The Analytic and the Synthetic: an Untenable Dualism," *John Dewey: Philosopher of Science and Freedom* (Ed. S. Hook, New York, Dial, 1950).
- WHITTEMORE, R. "The Metaphysics of the Seven Formulations of the Moral Argument," *Tulane Studies in Philosophy*, vol. 3 (Hague, Nijhoff, 1954).
- WITT-HANSEN, Johannes *En kritisk analyse af materiebegrebet hos Newton, Kant, og Einstein* [with English summary], (Copenhagen, 1958).
- WOLFF, Robert Paul "Kant's Debt to Hume via Beattie," *Journal of the History of Ideas*, vol. 21 (1960), pp. 117-123.
- *Kant's Theory of Mental Activity* (Harvard, 1963).
- ZOCHER, Rudolf "Kant's transzendente Deduktion der Kategorien," *Zeitschrift für philosophische Forschung*, vol. 8 (1954), pp. 161-194.
- "Zu Kants transzendentaler Deduktion der Ideen der reinen Vernunft," *Zeitschrift für philosophische Forschung*, vol. 12 (1958), pp. 43-74.
- *Kants Grundlehre* (Erlangen, 1959).

II. MORAL OBLIGATION

KURT BAIER*

A theory of obligation must contain at least three ingredients: an exposition of exactly what sorts of things are asserted in obligation claims; an explanation of exactly how one would go about confirming or disconfirming such claims; and lastly, in view of their important and serious practical implication, a justification of the practices in which the making and establishing of obligation claims is anchored.

I

Central to our inquiry are the claims in which obligations are ascribed to persons. Every such claim embodies what I shall call "directives." A directive is the content of speech acts capable of guiding those to whom it applies. To be suitable for use in directives, an expression must be capable of being taken as such a guide to action. It is so capable if it provides a model for judging action as in accordance with or as contrary to it. It must, in other words, formulate for someone a possible course of action. "Taking the firm's car" would satisfy this condition; "being six feet tall" or "being an eclipse of the moon" would not.

Begin with the simplest face-to-face contexts and the simplest kinds of directive in which only the minimum information is spelled out in the directive itself, and the rest supplied by the context. In such cases, I shall speak of:

1. PURE DIRECTIVES. My own position in discussing such pure directives is that of an observer describing, from a birds-eye point of view, situations in which speakers, hearers, and critics play the standard roles possible in connection with directives. Pure directives have only three dimensions, the "addresser," the "addressee," and the "assignment." In pure directives the addresser is never named or referred to but must be the person using the appropriate linguistic expression, thereby giving a directive and addressing it to the hearer, thus making him the addressee. The fact that the

addresser is addressing the directive to a certain person may be signalled simply by speaking in his direction or by naming him.

The context usually makes clear the relation between the addresser and the addressee. A person's characterization of the directive, and so its purported purpose and point, must to a considerable extent depend on his interpretation of that relation. Without a knowledge of this it is impossible to evaluate the directive. And in the absence of such evaluation one could not evaluate the performance of the addresser or the addressee. If the addressee does not stand to the addresser in a relation of subordination, either *de facto* (being held at gun point, or being under his thumb) or *de jure* (a subaltern, or a citizen) then the directive normally cannot be taken as a command or an order but should be taken as instruction, advice, a recipe, and the like. An important difference will be whether the assignment in the directive will have to be interpreted as "end-setting" or "end-promoting." In the former case, the addressee is directed to do something which, for whatever reason, the addressee wants him to do. In the latter case, the addressee is directed to do something which is held out to him as a way of attaining his own goal.

The status and point of the directive determines a suitable framework for judging its merit, i.e., its adequacy in relation to its status and point, i.e., its applicability and its soundness. What it purports to do, and for whom, settles the question of its applicability to the addressee. If "Give up bread and salami" purports to be an end-promoting directive addressed to a person who wants to lose weight, it is not applicable if it has been addressed to Jones who does not want or need to lose weight. The soundness of an applicable directive hinges on whether following it would serve the purpose determined both by its status and point, and by the peculiarities of the person to whom it applies.

Lastly, when we know the merit of the directive,

* Work on this paper was made possible by a research grant from the Carnegie Corporation of New York and the International Business Machines Corporation to the University of Pittsburgh to conduct a philosophical investigation of American values. I have greatly profited from discussion with my colleagues, especially N. Rescher and J. Schneewind.

we can assess the performance of the addresser in giving it, and of the addressee in following or ignoring it. This point is of particular importance for an understanding of obligation claims. For, as has often been noticed, they have a peculiarly strong and puzzling "binding force," as it is usually called. Thus, when we say that someone at a particular time had a moral obligation to stop the car, we imply that at that time a directive to stop the car was applicable to him and sound, and one with a peculiarly strong (moral) binding force. And this last point means that his failure to follow that directive would have to be judged particularly severely. I return to this point in Part IV of this paper.

2. **QUALIFIED DIRECTIVES.** If we now envisage more complex social situations than the simple face-to-face one of Sect. 1, we can see the advantage of incorporating into directives some of the information which otherwise has to be gleaned from the context. In the first place, the addressers of directives can thereby make clear what status and point they take their directive to have. They may say, "I would *advise* you to give up bread . . .," "I *promise* to be home early . . .," or "I *order* you to stop the car. . . ." The inclusion of these "indicator expressions," may but need not be decisive. A student saying to his teacher "Change my grade" cannot turn this directive into an order simply by inserting the indicator term, "I order you . . .," though a threatening tone and perhaps a gun may do the trick.

Secondly, the addresser can make explicit what he takes to be the conditions of applicability and soundness. He may, for instance, add, "*if you want to lose weight*, I'd advise you to . . .," or "*Unless you want to die of a heart attack*, you must. . . ." In these cases the if—and unless—clauses spell out what the addresser considers the conditions of applicability.

Thirdly, he may give greater specificity to the directive and so facilitate judgment of its soundness by clarifying the precise part he thinks the accomplishment of the assignment would play. Thus, he may say, "If you want to lose weight, you *must* . . ." or " . . . you *can* . . ." or " . . . you *might* . . ." thereby indicating whether he claims that the addressee's accomplishing the assignment is a necessary, a sufficient, or a helpful condition of his attaining his end.

He may also indicate how he, or how in his

estimate, others would assess the soundness of the directive and accordingly the merit of the addressee's performance in following or disregarding it. Words such as "ought," "should," "obliged," "obligated," "wrong," "immoral," "unlawful," can be incorporated in directives to make points of this sort. In all these cases, we can still speak of the person enunciating the directive as "giving" it. But we must remember that the role of the giver is very different when he "gives" a directive as an order or as a piece of advice, as a warning or a prayer, as a messenger transmitting someone else's orders, or as a housewife confiding a recipe to another. In many cases, the role is none too clear as in the case of a judgment of the Supreme Court, of a policeman telling someone he cannot park in a certain spot, or a father telling his son that it is wrong for him to date a girl of another race.

A specially important type are moral directives, that is, those which, other things being equal, it would be (*morally*) wrong for the addressee not to follow, or putting it in other words, those which, other things being equal, the person to whom they apply (*morally*) ought to follow. The questions which, as philosophers, concern us most in this connection are these: precisely what is the "binding force" of moral directives and to precisely what sort of directive is such a binding force rightly attributed?

3. **OBLIGATION CLAIMS.** Such claims clearly are directives of a certain sort. But what sort? If it is wrong for someone to do *x*, does he *ipso facto* have an obligation? If so, is it a moral obligation not to do *x*? And what is the relation between obligation and the so-called "moral ought"? If someone has, let us say, a legal (that is, a nonmoral) obligation to do *x*, does it follow that he morally, or that he non-morally, ought to do *x*?

We can most profitably answer these and other questions by a brief preliminary discussion of some points in H. L. A. Hart's theory of obligation, as stated in his essay "Legal and Moral Obligation," and in his book, *The Concept of Law*.¹ There he maintains that moral obligation covers only a *segment* of morality and that the wrongness of a certain sort of behavior does not entail that he has a moral obligation not to do it. For although killing and torturing are wrong, "It is absurd to speak of having a moral *duty* not to kill another human being, or an *obligation* not to torture a

¹ "Legal and Moral Obligation" in *Essays in Moral Philosophy*, ed. A. I. Melden (Seattle, University of Washington Press, 1958), pp. 82-107; and *The Concept of Law* (Oxford, Clarendon Press, 1961).

child."² Conversely he maintains that obligation is not necessarily or typically moral, any more than "ought," "must," and "should." In fact, he argues that obligation and duty are primarily at home, not in morality, but in "the legal world." It is correct to ascribe moral obligations only in special situations which exhibit striking resemblances with those in which we ascribe legal obligation. This, in Hart's view, is why there is no necessary connection either between the claims that it is wrong for *N* to do *x* and that *N* has an obligation not to do *x*, or between the latter and the claim that *N* morally ought not to do *x*.

Let us grant that it would often be inappropriate ("absurd" is surely too strong) to speak of an obligation not to torture children. But is the explanation of this that we can never derive a moral obligation not to do *x* from the fact that it would be morally wrong to do *x*. Surely not. If it is *morally wrong* to pass by when an injured motorist calls out for help, then in such a situation one has a *moral obligation* to give such help. Putting it more generally, we can say that whenever a wrong-claim indicates a *task* to someone, then it constitutes a moral directive and being the addressee of such a moral directive amounts to having a moral obligation. Why then does the wrong-claim, "It is wrong to torture children," not *seem* to give rise to an obligation? The answer is that, where the moral directive is negative, where it *forbids* rather than *bids* a person to do something, it may not indicate a task to the person referred to (or to anyone, for that matter). For that person may not have the opportunity, let alone the motivation, to do what the directive forbids him to do.

Since this is, one hopes, the case with most people where torturing children is concerned, it is odd, even insulting to say to someone that he has an obligation to refrain from torturing children, for it plainly implies that he would want to torture them if he had a chance. Yet, insulting though such a suggestion may be, it need not be absurd or even untrue. Sometimes a perverse desire, a need for information, or a so-called higher goal may be thought to require the torture of children. Then not torturing them may become a "self-disciplinary" moral task for such a person, and so a moral obligation. We tend to overlook such tasks because of the preponderance of what I shall call "productive" ones, and because most self-disciplinary tasks arise in conjunction with productive tasks. When

we have a productive task, one requiring us to do something, to bring something about or prevent something—say, aiding someone in distress or protecting someone in danger—we often have a self-disciplinary one as well, namely, overcoming our reluctance to engage in this productive task. And when we have a self-disciplinary task, we also usually can do something about preventing the recurrence of the temptation. The traditional association between obligation and effort of will, a commonplace in everyday moral thinking and, at any rate since Kant, in moral philosophy, would seem to show that obligations quite often involve self-disciplinary tasks. If a moral directive here sets such a self-disciplinary task, and so gives rise to an obligation, clearly the addressee morally ought to do what constitutes its discharge.

This brings us to the question of whether *all* obligation claims entail moral-ought claims, or whether only *moral* obligations claims do so. We shall be clearer about this question if we keep apart two ways of distinguishing between moral and nonmoral obligation claims. We may classify an obligation as moral if it is morally binding; if having the obligation entails that one morally ought to do what would constitute its discharge. I shall call this the *binding* sense. Or, we may classify an obligation as moral if it has come into existence in a certain way, namely, by someone's falling under an appropriate general moral directive and so having a moral task set for him. Jones's obligation to give assistance to the injured motorist by the roadside has come about in this way. He has come under the general moral directive, "Aid those in distress." Since this directive also imposes a moral task, it given rise to a moral obligation in this second sense. I shall call this the *genetic* sense. It distinguishes moral obligations from others, such as legal, religious, and social, on the basis of the corresponding kinds of directives which generate them. Jones's obligation to complete, each year, an income tax return by a certain date is a *legal* and not a *moral* obligation in the genetic sense. Of course, this leaves open the question of whether it is moral or not, in the binding sense. Now, whereas Jones's obligation not to work late on Friday, as he had *promised* his wife, is not moral but *promissory*, his obligation to keep promises is moral, not promissory (in the genetic sense). However, both these obligations, the more general and the more specific, are moral in the binding sense.

² Hart, "Legal and Moral Obligation," *ibid.*, p. 82.

The distinction just drawn helps us avoid the error of thinking that because there plainly are obligations which are nonmoral (in the genetic sense), there must be obligations with nonmoral (e.g., legal) binding force, i.e., obligations which are nonmoral (in the binding sense). I shall furthermore argue below that every obligation, even those which are nonmoral in the genetic sense, must be capable of reformulation in such a way as to make them moral also in the genetic sense. What is distinctive about such cases (e.g., obligations arising from promises) is that their genesis requires both a general moral directive "keep promises" and an empirical fact—e.g., the fact that Jones did promise not to work late—which defines the specific task in which the obligation consists.

The account of obligation which I offer is this: that obligations arise when and only when a morally binding directive gives rise, not simply to an assignment which may be effortlessly performed (as, for most of us, is the assignment not to torture children), but to a *task*, whether merely self-disciplinary or also productive. What distinguishes the different *types* of obligation (moral, legal, etc.) from one another is not the difference in the binding force (whose meaning and justification I discuss in Part IV), but rather the different genetic factors which transform an assignment into a task. Where the obligation is merely moral, and cannot be characterized also as promissory or legal, this factor is either the temptation of the addressee to do what the directive forbids (torture children) or the effort needed to perform the productive task which the directive enjoins (repair damage done). Where the obligation can be characterized as promissory or legal, there the task arises not merely from a moral assignment involving effort, but from a "blank" moral directive, "keep promises," "obey the law," along with some fact which gives that directive content and sets a specific task—that Jones promised not to work late, that the law requires tax returns, etc. What is common to all cases of obligation is a moral directive which has *somehow* given rise to a task. The different ways in which the task arises generate the different "types" of obligation.

I shall now try, in Parts II and III, to substantiate my claim that no obligation of any type arises unless a morally binding directive is involved.

II

How would we find out or make sure whether someone had an obligation to do a certain thing? The most popular theory of the origin of obligations is probably the Will Theory. According to it, an obligation by someone to do something is brought into existence by the appropriate expression of someone's wish, will, or intention that someone, whether himself or another, do that thing. Supposing that, "If your eye offend thee, pluck it out," is the appropriate expression of the appropriate will, then if Jones's eye offends him, he has an obligation to pluck it out. Typically, this theory lays it down that the expression of a will is appropriate only if the possessor of that will is able and ready to impose an effective sanction for disobedience on the person who is intended to do the thing. This feature gives the theory great initial plausibility since it appears to explain better than its most plausible rival, the peculiar "force" of an obligation claim. It appears to provide an explanation of the way in which an obligation binds or ties or obliges the person who has it.

Let us distinguish a strong and a weak version. According to the former, the expression of the appropriate will is a necessary and sufficient condition of the origination of an obligation. Moral obligation, on this view, is a sub-class brought about by some privileged will, usually that of God or society. On the weak version, the expression of the appropriate will is a sufficient but not a necessary condition³ of the origination of an obligation. The Will Theory can be attached to various modes of expressing the will. Some have favored The Command, others the Law or Social Rule, yet others The Promise. However, all of them have this much in common: (a) What brings someone's obligation into existence is the fact that he comes to be, in the appropriate way, the recipient of a "directive" to do that thing, (b) that this directive has come into being through the appropriate expression of someone's will or intention, and (c) that non-compliance with the directive is normally followed by the imposition of some form of "sanctions" which are normally effective. On this view, the import of an obligation claim is this: someone has given a directive which when ignored is normally followed by the imposition of the sanc-

³ One might also hold that this is a necessary but not a sufficient condition for the existence of some types of obligation. Von Wright, for example, seems to hold that, provided certain other conditions are also satisfied, while someone stands in a relationship under "obligation norm" to someone, he has an obligation. Cf. e.g., G. H. von Wright, *Norm and Action* (London, Routledge & Kegan Paul, 1963), pp. 116 ff., and also his *Varieties of Goodness* (New York, The Humanities Press, 1963), p. 170.

tion and so the addressee is liable to incur the sanction. The method for finding out whether someone has an obligation to do a particular thing is thus simply to find out whether his case comes under an appropriate directive. I shall argue that the Will Theory requires us to by-pass the crucial question about obligation claims, namely, whether it would be wrong to ignore a directive of that origin.

The classical statement of a strong version of the Will Theory is to be found in John Austin's *The Province of Jurisprudence Determined*.⁴ Austin there claims that "duty" (which he uses interchangeably with "obligation") and "command" are correlative terms: "wherever a duty lies, a command has been signified; and whenever a command is signified, a duty is imposed."⁵ Thus to say that someone has an obligation to help others in need would (on Austin's theory) be to say that he has been commanded, i.e., told by someone able and ready to impose effective sanctions, to help others in need.

Consider the following argument (A):

- (1) Smith, a person able and ready to impose effective sanctions for disobedience, peremptorily says to Jones, "Do x ."
- (2) Hence Smith addresses to Jones a command to do x .
- (3) Hence Jones has an obligation to do x .

Whatever, within our framework, may be the exact relationship of superiority and inferiority between Smith and Jones on which the legitimacy of the step from (1) to (2) depends, it cannot depend on Smith's having the right to give commands, for the gunman who peremptorily says "Hands up" to the traveler he is holding up is commanding or ordering him to raise his hands, even though he has no right to do so. But just because he has no right to do so, the traveler has no obligation to obey. It is not merely that he has no moral obligation to obey, he has no obligation. The step from (2) to (3) hinges on the general proposition.

- (4) One has an obligation to obey sanction-backed commands addressed to one.

It is not plausible to say, even at first sight, that this is tautologous or otherwise sound. If, however, we attempt to derive (3) from (2) without relying on (4), then (3) must lack the force of an ordinary obligation statement. For what follows is at most

that since Smith is able and ready to impose effective sanctions on Jones if he disobeys, it will often be likely or perhaps certain that Jones will incur the evil of the sanction if he disobeys and so Jones will often have a good or conclusive reason for obeying. But, of course, sometimes he may have good or conclusive reason for thinking that he will not incur the sanction, and then he may have no reason whatever or not a very strong reason for obeying. Thus, the conclusion, (3), does not render the meaning of "obligation." For in having an obligation to do x , one necessarily has a good reason for doing x , whereas in having been commanded to do it, one need not.

Even where having been commanded to do x yields a reason for doing something, it cannot be the same sort as that yielded by an obligation, for it does not yield a decisive reason against not doing so. Someone may try to justify or excuse the enormities he has committed on the grounds that he was commanded to perform them, and that he was threatened with very serious sanctions if he refused, and that he thought it very likely that the sanctions would be imposed, and that he was therefore compelled (or obliged) to perform them. This really may excuse what he did, though it would hardly justify it. However, if, at great peril to himself, he disobeyed the command, say, to beat a fellow prisoner to death, the fact that he was commanded and therefore had a reason to do so, does not establish that he was not justified in refusing to do it. By contrast, a man's having an obligation to do something does imply that he would not be justified in not doing it. The point of this example is to draw attention to the following difference: " N has an obligation to do x " entails that " N is justified in doing x and not justified in not doing x ." By contrast, " N was commanded to do x " does not entail that N is justified in doing x , or that he is not justified in not doing x . It entails only that N has a good reason for doing x , and so that N usually has an excuse for doing x , if doing x is the sort of thing which requires an excuse. But this is compatible with his having a decisive reason against doing it.

The subtlest among the defenders of the Will Theory is H. L. A. Hart. Against the Command version he argues that a person's having been ordered to do something is neither a sufficient nor a necessary condition of his having an obligation to do it. He offers two major modifications designed

⁴ John Austin, *The Province of Jurisprudence Determined and the Uses of the Study of Jurisprudence*. Reprint edition, Library of Ideas (London, Weidenfeld and Nicolson, 1954).

⁵ *Ibid.*, p. 14.

to overcome the shortcomings of the Command version while retaining the essentials of the Will Theory.⁶

The first modification is the replacement of The Command as an obligation-creating social device by another, the Social Rules of Obligation.⁷ Falling under such a rule is, on his view, a sufficient (and a necessary?) condition of having an obligation to do as the rule requires, provided only one further condition, namely that the rules have an "internal aspect," is also satisfied. The introduction of this further condition is Hart's second and most original contribution, about which I shall say more below.

What, then, is this more suitable obligation-creating device? It is a certain subclass of social rules which in turn is a certain subclass of social regularities of behavior. According to Hart, a social rule requiring the doing of x differs in three respects from a mere social habit of doing x . In the former case, but not in the latter, (i) deviations from the regular behavior "are generally regarded as lapses or faults open to criticism, and threatened deviations meet with pressure for conformity"; (ii) "deviation from standard is generally accepted as a *good reason* for making (such criticism)"; and (iii) "some at least must look upon the behavior in question as a general standard to be followed by the group as a whole."⁸ Thus social rules are what I call general *directives*. Of course, not all social rules are rules of obligation. Rules of etiquette or correct speech, for instance, are not.⁹ Social rules of obligation are that subclass of social rules which are characterized by the following three features: (iv) they are thought important and therefore "the general demand for conformity is insistent and the social pressure brought to bear upon those who deviate or threaten to deviate is great"; (v) they are thought important because they are "believed necessary to the maintenance of social life"; (vi) they are "those characteristically involving sacrifice and renunciation."¹⁰

Leaving aside, for the moment, Hart's most original contribution which I have not yet examined, Hart's method of substantiating an obligation claim can perhaps be sketched as follows in argument (B):

- (1) The rule-formula, R ("Do x ") satisfies conditions (i)-(vi).
- (2) So there is a social rule of *obligation*, R ("Do x ").
- (3) And N 's case falls under R .
- (4) So N has an obligation to do x .
- (5) So N ought, other things being equal, to do x .

It is clear that on this model, as on those of the Command, the soundness of the argument hinges on the acceptability of a general proposition, in this case,

- (6) One has an obligation to act in accordance with rule-formulae satisfying conditions (i)-(vi).

Calling a formula which satisfies these conditions a social rule of *obligation* may conceal the need for (6), but it cannot eliminate it. If Hart's position strikes us as an advance on the Command Theory, this must be due to the fact that (B-6) appears to be more acceptable than (A-4).¹¹ Wherein, then, does this appearance of greater acceptability lie? It seems to me to lie in the fact that social rules appear to be a less "naked" imposition of someone's will on that of another. For the content of a person's obligation arising out of a social rule is in two respects less dependent on the brute will of another than an obligation arising out of a command. A rule is necessarily, or at least typically, a general directive, where a command may be, and typically is, particular: a rule is meant for more than one occasion and for more than one person. This naturally leads to rule-formulations which are general, and so rules *apply* to people satisfying the conditions specified in the rule, rather than being *addressed* to people irrespective of the conditions they satisfy. Thus, rules are less will-dependent, less arbitrary, than commands, in respect of *the persons of whom* a certain sort of conduct is required. In respect of *what* is required, however, they are quite as will-dependent and arbitrary as commands. They are moreover, more "naked" than promises in this respect: that whereas a person voluntarily comes under a promise, since he gives it himself, he does not voluntarily come under a command or social rule. Hence the will-dependence of what is

⁶ Hart, *The Concept of Law*, op. cit., p. 81.

⁷ *Ibid.*, especially p. 83.

⁸ *Ibid.*, p. 55.

⁹ *Ibid.*, p. 83.

¹⁰ *Ibid.*, pp. 84-85. (I mention in passing that clearly Hart's account does not tell us correctly when "rules are conceived and spoken of as imposing obligations," for the rules which impose, e.g., social obligations have few if any of his three features (iv)-(vi).)

¹¹ This was mentioned earlier.

required by a promise-created obligation is not in itself objectionable, whereas that of a commanded or rule-created one is. A special justification of a social device creating such obligations is needed.

Hart's stress on "the internal aspect" of social rules is designed, I think, to by-pass the need for their justification. For a social rule to exist at all, there must be, Hart claims, an internal aspect of it.¹² That is to say, at least *some* members of the society whose rule it is must look upon it from the internal point of view, i.e., must regard the behavior required by it as a general standard to be followed by the group as a whole, and therefore deviation from that standard as a good reason for criticizing deviants.¹³ Obligation claims are thus made *from* the internal point of view. Those who look upon a rule from the external point of view "will need for its expression, 'I was obliged to do it,' 'I am likely to suffer for it if . . .,' 'You will probably suffer for it if . . .,' 'They will do that to you if . . .,' But they will not need forms of expression like 'I had an obligation' or 'You have an obligation' . . ."¹⁴ Thus, Hart's second condition ensures that those who *use* obligation language are also necessarily those who "accept," i.e., support and maintain, the rules which give rise to such obligations, and regard deviation from those rules as good reasons for criticizing deviants. In a sense, such obligations are therefore self-imposed. Hart's second condition thus transforms the obligation-creating device into one which, in respect to "naked imposition," lies somewhere between the Promise and the Command. The will expressed in social rules of obligation is not exactly the obligated person's own will (as it is in a promise), but neither is it exactly that of another (as it is in a command).

All the same, even Hart's modified version of the Will Theory is untenable, for he fails to give a satisfactory answer to the question of whether or not those who look upon the rules from the external point of view have obligations. Clearly, we cannot say that those who look upon the rules from the external point of view have obligations. For Hart's account suggests that having an obligation requires acknowledging, at least to oneself, that one has an obligation. If we waive this requirement,

Hart's position becomes much the same as Austin's. For since these "outsiders" do not *accept* the rules, they think of them merely as obstacles in their way and will acknowledge at most that their existence *obliges* them to act as required. But Hart rightly rejects this as an account of obligation.¹⁵

But to say that these "outsiders" do *not* have obligations is also unacceptable, for it creates a wholly artificial distinction between the "insiders" and the "outsiders." Suppose Smith and Jones are aliens. The law under which they live requires aliens to notify the police of any change of address within a week. Both have moved recently. Then Jones who "accepts" the law has an obligation to report the change, but Smith who does not "accept" that law, but merely wishes to keep out of trouble, does not have an obligation to report. To accept this distinction would be to give the wrong meaning to obligation-claims since they are, in a certain sense, "categorical": one cannot refute an obligation claim by showing that the addressee's discharge of his obligation would not be a means to any end of his. But on the view under scrutiny, that is precisely what we are allowing such an "outsider" to do. When Jones says to Smith, "We have an obligation to notify the police," Smith can reply, "You have, since you accept the law. I don't for I do not accept the law. I may be obliged to do it to avoid the fine, but I do not have an obligation to do it." And if he knows he can get away with not doing it, then *he* has *no reason* for doing it.

There is only one other possibility, namely, to say that the question of whether such "outsiders" have an obligation to obey the rule simply does not arise. This seems to be Hart's own view. On that interpretation, the very question of whether a particular person has an obligation to do a thing arises only for those "insiders" who live under a social system whose rules of obligation they accept. They can conclusively answer it by finding out whether that person's case comes under such a rule. For the others it does not arise and so need not and cannot be answered. On the question of whether someone has an obligation, "insiders" and "outsiders" must inevitably be at cross purposes. There is no "neutral" or "objective" standpoint from which the question of whether someone "really"

¹² Hart, *The Concept of Law*, *op. cit.*, especially pp. 55-56.

¹³ *Ibid.*, p. 88.

¹⁴ *Ibid.*

¹⁵ It is possible that although he rejects it as a general account of obligation, he might accept it as long as it applies only to the minority who adopts the external point of view. I think it can be shown that such a view is unsatisfactory since it comes up against similar objections to those made against Austin, namely, that there would then seem no good reason why any "outsider" should bother with his obligations as long as he can avoid the sanction, and does not mind the disapproval of his fellows.

has an obligation to do x can be raised and answered. The only "objective" point that can be established is whether a person's case comes under a rule, and whether he will go on to say that he has an obligation, but not whether he is right or wrong in going on the way he does.

Hart's account of obligation is thus a sophisticated Attitudinal (Emotive) Theory. He explains the peculiarities of obligation claims as a combination of two quite different ingredients. The first is a fairly straightforward empirical, sociological one, namely, the ascertainable fact that a person's case comes under a rule of obligation which is actually "accepted" by a significant portion of his group, regardless of the nature and cogency of their reasons for accepting it.¹⁶ The second is tied to the speaker's "acceptance" of the rule, from the "internal point of view."

My objection to Hart's theory can be put in the form of a dilemma: Can "outsiders" have obligations or can they not? If we say they cannot, we have plainly failed to give a correct account of obligation claims. For such claims involve directives purporting to be "unconditionally" applicable and sound, and possessing a peculiarly strong binding force. But if "outsiders" cannot have obligations, then the applicability of obligation claims is conditional on a person's actually "accepting" a given rule.¹⁷

If we wished to convince an "outsider" that he is obligated to do a certain thing, we should have to argue, not that he *actually had* this obligation but, absurdly, that he ought to accept the obligation rules of the group so that he *can* have obligations. In fact, of course, it is sufficient to provide adequate reasons in support of the obligation rule. For when such reasons are available, nonacceptance of the rule is irrelevant since in that case it has been shown that the rule *ought* to be accepted.

Embracing the other horn of the dilemma, saying that "outsiders" also have obligations, reduces Hart's position to that of J. Austin.¹⁸ For then, since the directives embodied in these obligation rules are not regarded as requiring justification but merely as being widely accepted, they are brute

compulsions for those who do not happen to accept them.

One last point. Hart's main concern in giving a "content-independent account" of legal obligation is to avoid landing in the camp of the Natural Law theorists and so having to maintain that iniquitous social rules cannot be law.¹⁹ No such consequence need follow from the *rejection* of Hart's view that obligation-claims (apart from drawing attention to the fact that a person comes under a rule of obligation) merely express a certain attitude of "acceptance" toward these rules. To show this, let us consider argument (C):

- (1) One has a (moral) obligation to obey the law as such.
- (2) R ("Do x ") is a valid law in society S .
- (3) So *there is a presumption* that people living in S have a (legal) obligation to obey R .
- (4) But Jones lives in S and his case comes under R .
- (5) So *there is a presumption* that Jones has a (legal) obligation to obey R .
- (6) So, other things being equal, Jones (morally) ought to obey R .

We avoid the consequence to which Natural Law theory is committed by insisting on the presumptive nature of the steps from (2) to (3) and from (4) to (5). Laws with a certain content may not give rise to obligations. But provided the content of R is not objectionable on moral grounds, the presumption that Jones has a (legal) obligation to obey R holds good. Hence there is no incompatibility between the claim (A), which Natural Law theorists rightly insist on, that unless (1) is sound, no valid law can give rise to obligations, (B), which I insist on, that some valid laws do not give rise to (genetic) legal, hence not to (binding) moral obligations, and (C), which positivistically inclined philosophers, like Hart, insist on, that we are justified in asserting that, unless the contrary is proved, the fact that since a certain valid law *enjoins* the doing of x , those whose case comes under it may be thought and said to *have a (genetic) legal obligation* to do x .

Perhaps the most plausible version of the Will

¹⁶ Hart's explanation of "acceptance" is in terms of how people regard deviations from the rule. In *The Concept of Law* he speaks indifferently of their regarding it as a *signal* (p. 87), or a *reason* (p. 88), or a *justification* (pp. 54-55) for a hostile reaction to the rule-deviants, yet surely whereas a signal does not purport to have any cogency, a justification purports to have a very high degree. The second and characteristic ingredient (pp. 86, 88) is an attitudinal one, namely, that only "insiders," only those who "accept" the rule in question, will actually use *obligation* language (thereby expressing their acceptance of the rule), though they may apply it equally to "outsiders."

¹⁷ For further details on this point, see Pt. III, sect. 3 below.

¹⁸ Cf. Pt. II above.

¹⁹ Cf. Hart, *op. cit.*, pp. 206-207.

Theory is that which employs as its obligation-creating device The Promise. Here, the giver and the addressee of the directive are one and the same person. The person to whom the promise (not the directive) is given, thereby receives the right to set in motion, if necessary, whatever machinery there is for exacting fulfillment of the promise. The inference from the nonmoral premiss, that someone in certain circumstances uttered certain words constituting a direction to himself, to the conclusion that he morally ought to follow the directive, seems wholly licensed by what is necessarily involved in a promise, and so to be analytic. All the same, the validity of the inference plainly depends on *two sorts of* assumptions, assumptions concerning the inferences licensed by the social institution itself, and assumptions concerning the power of a social institution to license such steps. If the move from "Jones promised to do x " to "Jones has an obligation to do x " is defeasible, and if the defeating condition is itself moral in nature, say, "Immoral promises do not give rise to obligations," then the social institution of promising is an obligation-creating device which rests on moral convictions. Hence in the case of some social institutions, say, child-marriage vows or slave labor contracts, which rest on no moral convictions, these "promises" do not give rise to genuine obligations. We should therefore admit that what gives rise to moral obligations in the case of promises is not saying, according to the rules of the institution, "I promise," but the fact, where it is fact, that it is wrong to break our promises.

In view of the extensive discussions this topic has recently received, it would be tedious to go over the ground once more. I therefore simply assume that the Will Theory fails even in the case of the most promising candidate, The Promise.²⁰

I conclude that even the most persuasive versions of the Will Theory are untenable. Their chief weakness lies in their attempt to derive obligations analytically from the expression of some will. The conclusions so derived have inevitably been shorn of all normative content. Hence, there is a tendency to smuggle in surreptitiously a supplementary principle such as that a will so expressed (promise, accepted social rule, law) should be regarded as

binding, or that a will so expressing directives has authority to bind, or that the will of such a person is certain to give directives which have a claim to be regarded as binding. But in the absence of such supplementary principles, the will models employed produce at best directives which it may be excusable, perhaps wise (but not obligatory), to follow.

III

Another group of theories, which I call teleological theories, attempt to derive obligation-claims from the conclusions of some form of practical reasoning. Consider a simple case. Jones is an adipose man with a certain heart condition which requires that he lose weight. Someone offers the following piece of advice: stop eating bread and salami. He might back up his advice as follows in argument (D):

- (1) If anyone with a certain sort of heart condition and suffering from adiposity is to minimize the risk of a heart attack, he must lose weight.
- (2) If anyone who suffers from this condition and also regularly eats bread and salami is to lose weight, he must stop eating bread and salami.
- (3) But you have a heart condition of this sort, you suffer from adiposity, and you regularly eat things, and it is your end to minimize the risk of a heart attack.
- (4) So, other things being equal, you must stop eating bread and salami.

Such arguments are closely parallel to those we noted above, in connection with obligation-creating social devices, such as commands and rules. Here, too, we find general directives stating or implying applicability conditions [i.e., (1) and (2)]; assertions to the effect that these conditions of applicability are satisfied in the case of a particular person [i.e., (3)]; and a conclusion which applies the directive, that is, the consequent of (1) and (2), to that particular person. We must, however, note five important points of difference.

(a) The major premisses contain not only a directive, but also a statement of a "connection in

²⁰ For a fuller discussion of both the assumptions I called in question, cf., John Searle, "How to Derive 'Ought' From 'Is'," *The Philosophical Review*, vol. 73 (1964), pp. 43-58; Roger Montague, "Ought' From 'Is'," *Australasian Journal of Philosophy*, vol. 43 (1965), pp. 144-167; Evan K. Jobe, "On Deriving 'Ought' From 'Is'," *Analysis*, vol. 25 (1965), pp. 179-181; W. D. Hudson, "The 'Is-Ought' Controversy," *Analysis*, vol. 25 (1965), pp. 191-195; Antony Flew, "On Not Deriving 'Ought' From 'Is'," *Analysis*, vol. 24 (1964), pp. 25-32; James E. McClellan and B. Paul Komisar, "On Deriving 'Ought' From 'Is'," *Analysis*, vol. 24 (1964), pp. 32-37; James and Judith Thomson, "How Not to Derive 'Ought' From 'Is'," *The Philosophical Review*, vol. 73 (1964), pp. 512-516.

nature." This connection may be of various kinds, and assertions to the effect that it holds can be true or false. Thus, if a heart attack is bound to occur unless the agent loses weight or remains completely motionless and if the latter is impracticable, then his losing weight is a *necessary condition* of the non-occurrence of a heart attack. In that case, we can say that to avoid a heart attack the agent *must* lose weight. If it is possible for him to remain motionless, then losing weight is not a necessary but at best a sufficient condition of the non-occurrence of a heart attack. If it is at any rate a sufficient condition of reducing the chances of a heart attack, then we can say that to avoid (or reduce the risk of) a heart attack, he *can* lose weight. If losing weight is thought greatly preferable to lying still, this can be indicated by saying that he would be *well-advised* or that he *should* lose weight.

This sort of argument is designed to support an end-promoting directive which is asked for and given when a person has a certain end and does not know how to attain it. It is assumed that when the directive is given, so is a solution to the questioner's problem. Hence, the conclusion which can be formulated either in terms of an imperative or in terms of words such as "can," "must," "should," or "ought," is felt to be a *directive* even when it is not in imperative form. Conversely, even when formulated in the imperative, the conclusion is felt to be capable of being true or false, for it implies that there is a connection in nature between the behavior delineated in the conclusion, and the end attributed to the agent in the minor premiss. The only difference between the imperative and the other formulations is that the former says nothing whatever about the nature of this connection, and so indicates nothing about the merits of the solution or the chances of success in following it.

(c) The conclusion must therefore be read in two ways, as an end-promoting directive to someone seeking the solution of a problem, and as *descriptive* of a certain sort of connection in nature, namely, a possible means-end connection. Our conclusion, "Stop eating . . ." or "You must stop eating . . ." is interpreted as a directive if, supposing there to be a discrepancy between the directive and Jones's behavior; it is *he* or *his behavior* that is *criticized* on that ground. It is read as descriptive if supposing there to be no such discrepancy and Jones still does not attain his end, namely, to lose weight, it is *the conclusion* which is criticized on that ground. Of course, just how soon he must lose weight and how much and under what conditions

depends on the degree of explicitness with which the means-end connection is stated in the conclusion. The fact that the conclusion can be taken descriptively at all shows that the connection in nature, asserted in the major premiss, must be taken as the *ground* of the conclusion when taken as a directive, and that the merit of the directive is therefore based on whether or not that connection holds.

The main conclusions I wish to derive from this are two: (i) that the reason why it is right to interpret the remarks "Stop eating . . ." or "You *must* (should) stop . . ." as directives is that they are offered in a context in which the person addressed is assumed to have a certain end to which the action delineated is implied to be a means. Neither the grammatical form nor the occurrence of the words "must," "should," or "ought" are the reasons for it, since the substitution for these words of the word "can," which clearly has no directive force by itself, does not eliminate the directive force of the remark as a whole. (ii) That the function of the words "can," "must," and "should" in such remarks is rather to indicate the precise nature of the connection between the action recommended and the end of the person addressed.

(d) We can now state one important difference between two ways of interpreting a directive such as, "If you cannot stay in bed, stop eating bread and salami." Taken as an order or command, it is an end-setting directive: the if-clause specifying the condition under which the directive is to be followed. In that case, the question of its soundness cannot arise, but that of its authorization can. But taken as a piece of advice, it is an end-promoting directive, the if-clause stating the applicability-condition. In that case the question of its authorization cannot arise, but that of its soundness can. Of course, a remark may have to be taken in both ways, e.g., a doctor's order by an army doctor to a sick recruit.

(e) As in the case of commands and rules, we do not find the word "obligation" in the conclusions of practical arguments of this form. To derive obligation-claims, we should therefore have to introduce an additional premiss, such as (5) or (6) in argument (D):

- (5) One has an obligation to do what one must, can, and should do in order to attain one's end.
- (6) One has an obligation to do what someone (soundly) advises one to do in order to attain one's end.

Plainly it would be even harder to establish such a general proposition than it would be to establish the corresponding (A-4), (B-6), or (C-6).²¹

However, at least the words "must" and "ought" naturally occur in the conclusions of such arguments. It is therefore tempting to identify the conclusions of such practical arguments with obligation-claims. Thus von Wright says, "To show why something is an *obligation* founded on interest, is not to show that it is something we ('really,' 'innermost') *want to do*, but that it is something we *have to do* for the sake of that which we *want* (to be, to do, to have, to happen)."²² Von Wright thus implies that when we *have* to do something for the sake of some end, we have *some* kind of obligation to do that thing. It is the fact that we *have to do* something for the sake of that which we *want* (to be, to do, to have, to happen) which makes this an obligation; not, as Hume wrongly thought, the fact that there is something we really, innermost want to do. However, this fact simply does not make a line of action obligatory. It does not even necessarily make it something we ought to do. As Hart points out,²³ in this sort of situation we may be *obliged* to do this thing, but we are not obligated to do it.²⁴

To return to our argument (D), even the word "ought" in the conclusion introduces an important new element. For to say, in a context such as we have examined, that Jones *ought* to stop eating bread and salami, is to imply not merely that the action recommended is the best way of attaining a certain possible end, but also that it is an end which the agent would be well advised to adopt. If the agent, realizing that unless he runs he will miss the train, says to himself, "I must run," he leaves open the question of whether or not he is well advised to pursue the end he has, namely, to catch that train. If he says, "I *ought* to run," he implies that he would be well-advised to pursue that end.

Thus, whereas the occurrence of words such as "can," "must" and even "should," does not imply

anything about the advisability of the pursuit of alternative courses of action to alternative ends, the occurrence of the word "ought" does have such implications. But even claims made by means of "ought" do not necessarily ascribe obligations. The main task for Teleological Theories is therefore to provide support for propositions such as (D) (5) and (6). I briefly examine three popular theories to expose the essential weaknesses of this model.

(III-1) RATIONAL EGOISM, interpreted as a theory of obligation, maintains that one has an obligation to do whatever and only what one ought to do, and that there is only one thing one ought to do, namely, to pursue those ends the attainment of which would be in one's best interest. Although this theory has not been held by any great philosopher, a brief discussion of it will help to make clearer the strengths and weaknesses of the more sophisticated theories in this group. All theories of this group make two important moves. The first is to distinguish between those ends or goals which a person *finds himself* having and those which he acquired as a result of deliberation. Standing back from our own past actions, noting the circumstances under which we have come to have, and then to pursue and sometimes to reach certain ends, we find that quite frequently we afterwards regretted having pursued and reached some of these ends, because of their bad consequences or the loss of the better things we might have had instead. Egoism recommends that we work out for ourselves in the light of our knowledge, our predilections, preferences, likes and dislikes, our capacities, talents, energies, and skills, our opportunities and resources, a life plan whose realization would make our life as rich and worthwhile as possible, and that we then plan the steps necessary for its realization. Of course, as we grow older and wiser, we may have to modify this plan in the light of our changed insights or the changed circumstances. But at any given time we should, with the aid of such a plan, be able to judge not only what will be the best means to an end we then find ourselves

²¹ Cf. n. 3 above.

²² *Varieties of Goodness*, *op. cit.*, p. 170.

²³ Hart, *The Concept of Law*, *op. cit.*, pp. 80-81.

²⁴ In his interesting book, *Practical Reasoning* (Oxford, Clarendon Press, 1963), chap. XII, David P. Gauthier, after showing in some detail, and in my opinion quite correctly, what it is to be obliged to do something then, wrongly, explains having an obligation as a special case of being obliged, namely, being obliged to someone *by* some "obliging factor" (p. 184). I suspect that Gauthier was misled by a failure to distinguish between "being obliged *by* someone to do something," which is not a case of having an obligation to do it, and "being obliged to someone *for* something" which is such a case. Because of this failure, he did not see that not only is having an obligation not a special case of being obliged (by someone to do something), but being obliged (to someone for something) is a very special case of having an obligation. Hence being obliged cannot be used to explain having an obligation.

having, but also whether the pursuit and attainment of such an end fits best into our life plan, or whether instead we should refrain from pursuing that end and pursue another more suitable end instead. At any given time what we ought to do is what is the best way of attaining those ends, the attainment of which best realizes a life plan whose realization will make our life as rewarding and meaningful as possible.

The second move is to offer an account of what makes a fact a reason for doing a certain thing. In any system of practical reasoning it is important that we be told not only *how* to reason with what facts, but also *why we should* reason in just that way, which will often require us to resist inclination. Such a method of reasoning must have some *appeal* to, or attraction for, the reasoner or else it will not be used. Rational Egoism offers a very attractive method, so attractive that it is never challenged, but serves rather as a paradigm, perhaps the only paradigm of practical reasoning. On this model, the criterion of a reason for doing something is that following it is in the agent's own best interest or for his own greatest good. Thus, a fact, *F*, is properly a reason for *N* to do *x*, if *N*, on being apprized of *F* and therefore doing *x*, is thereby promoting his best interest. Spelled out, this means, as we have seen, that in acting because of *F*, he is promoting an end, the attainment of which helps to realize a life plan which will make his life as rewarding and meaningful as possible. The model thus provides an extremely strong *justification* for this type of practical reasoning, resting on the acknowledged superiority of this mode of action over action on impulse. It may be virtually impossible to establish whether a given fact is a good reason for doing something, in this system, but it is hardly possible to challenge the framework. Insofar as we are willing to curb our impulses at all, are ready at all to refrain from pursuing ends we find ourselves having, we shall be open to the consideration that the pursuit of other goals would make for a richer, more rewarding, more worthwhile life. We should regard a person as worse than quaint if he seriously asked why he should do what is in his best interest.

It must be granted that Rational Egoism can give a plausible answer to questions of the form, "What ought I to do"? For it can answer not merely questions about what is a possible, the only possible, or the best way of attaining an end some-

one *happens* to have, but also to the question of which of the many ends he might pursue would be *the best one* to pursue. When we say "You ought to have been there, it was unbelievable" or "you ought to invest in Fixed Trusts" or "You ought not to take on any more speaking engagements," we may support these claims in just the way Rational Egoism indicates. Yet clearly these are not cases of someone's having an obligation. For if someone has an obligation to do something, then it follows that it would be wrong for him not to do it, but that does not seem to follow from these remarks.

(III-2) UTILITARIANISM²⁵ is similar to Rational Egoism, for it uses the same ingredients, but it serves them up in a different mixture. Act-Utilitarians say that one ought to pursue those ends whose attainment would be for the greatest good of the greatest number. An individual, therefore, has to consider not only his own life plan and what is necessary to realize it, but also the life plan of everyone else. Reasoning under this system, thus, also requires him not to pursue those of his ends whose attainment, though for his own greatest good, is not for the greatest good of the greatest number. Act Utilitarianism can claim to be much closer to our everyday moral convictions than Rational Egoism. We believe that morality requires us often to do things which are not in our best interest and to refrain from doing things which would be in our best interest. Rational Egoism flies in the face of this deep-seated conviction. Egoism is moreover necessarily useless as a way of deciding which of two people ought to refrain from pursuing an end which is his interest, in circumstances in which it is impossible for both of them to attain their ends. In these cases, Act Utilitarianism can sometimes yield a decision.

However, Act Utilitarianism has a much weaker answer to the question why we should use its method of reasoning. For to the question why they should prefer the greatest good of the greatest number to their own greatest good, Act Utilitarians must reply either that people *ought* to aim at the greatest good of the greatest number or that *they in fact* do. The former falls wholly outside their own conceptual framework and reduces Utilitarianism to Intuitionism, i.e., to the absence of any theory of obligation. The latter is often false, and so reduces the applicability of obligation-claims to

²⁵ Cf., e.g., J. J. C. Smart, *An Outline of a System of Utilitarian Ethics* (Melbourne, Melbourne University Press, 1961); R. B. Brandt, "Toward a Credible Form of Utilitarianism" in *Morality and the Language of Conduct*, ed. Hector Neri Castaneda and George Nakhnikian (Detroit, Wayne State University Press, 1963).

those who are psychologically so made that they do in fact pursue these ends. Like Hart's "insiders," Utilitarians must admit that those with a different psychological make-up *need* not, perhaps *ought* not to use its system of practical reasoning. On this view, too, morality is the system of reasoning used by and applicable to only those "insiders" who in fact have the utilitarian end. On this theory, a morality is much like the code of some coterie or religion.

From our point of view, the only difference between Act and Rule Utilitarianism is that instead of the pursuit of ends, the latter speaks of acting in accordance with rules. Its formula for determining one's obligations is: One ought to enter on that course which, if made a general rule, would promote at least as great a good of at least as great a number as any other course open to him, if made universal. This modification of Utilitarianism may well yield results still closer to our every-day moral convictions, but it is otherwise open to the same objections as Act-Utilitarianism.

(III-3) THE BASIC WEAKNESS OF ALL TELEOLOGICAL THEORIES WELL FORMULATED BY KANT. As he points out, all such practical reasoning is based on "hypothetical imperatives" and so lacks the peculiar force of obligation-claims which are categorical. Following Kant, one might try to explain the force of obligation-claims as a certain property of some imperatives, namely, being in a certain sense *unconditional*, such that "we cannot be free from the precept if we give up the purpose." The precepts which apply to a person do so irrespective of whether or not he has the end to whose attainment the precept is a means. But though important, possession of this feature is only a necessary, not a sufficient condition. There are directives which satisfy it without being obligations, such as the doctor's order to our adipose man. That order may well be shown to be applicable to Jones under the conditions outlined; its applicability is not conditional upon his *in fact* wanting to reduce: he ought to have this end whether or not he in fact has it; and it may be sound. Jones cannot be "free from" the precept if he gives up the end. Nevertheless, this is not an obligation. For surely despite all this, Jones is *entitled*, has a *right*, to go on eating bread and salami if he wants to. But he cannot have both an obligation to stop eating bread and salami and also a right to go on doing so if he wants to.

We need to distinguish two types of universal

²² To use Falk's useful term. Cf. W. D. Falk, "Morality, Self, and Others" in *Morality and the Language of Conduct*, *ibid.*, p. 43.

applicability and soundness: (i) where it is the case either in fact or of logical necessity, that all persons have the end, e.g., "You ought to do this if you value your happiness"; (ii) where the end in question is one which some persons sometimes do not have, but which any person always ought to have, e.g., "You ought to do this if you value your health." In case (i) a person would be free from the precept by ceasing to have the purpose and so Kant's unconditionality criterion rightly disposes in this case. But in case (ii) he cannot be free by ceasing to have the purpose and so that criterion by itself would, wrongly, uphold this case as an obligation.

We can represent the argument (E) of such a crude "Kantian" theory in this way:

- (1) One has an obligation to act in accordance with certain precepts, namely, those delineating actions which are means to ends one ought to have, whether or not one actually has them;
- (2) that is, ends from which one cannot be free even if one could give up the purpose which they serve;
- (3) that is, ends which are unconditional, and in a certain sense universally applicable and sound.
- (4) But health is such an end, "an ought through and through."²³
- (5) So precepts delineating actions which are means to the preservation of one's health give rise to obligations to follow them.

Now, a strong case can really be made for (4): because there is no single end or combination of ends (except perhaps their totality) whose abandonment would make the maintenance of one's health unnecessary or useless. There is thus a perfectly good sense (though not the sense which Kant has in mind), in which the maxim "preserve your health" is always applicable and sound, hence *unconditional*, hence categorical. All the same, the fact that an argument for doing what preserves one's health is conclusive, does not make doing it obligatory.

IV

Our brief review of some theories of obligation exposes their failure to explain and justify, in many cases even to attend to, the peculiar and seemingly obnoxious binding force attaching to obligation

claims. This conceals the justifiability of challenging those who make obligation claims until they can demonstrate their bindingness. But the received theories cannot meet such challenges. In answer to the claim that a directive issuing from some Authorized Will could not amount to an obligation unless it were wrong not to follow directives from such a source, the Will Theory can at most say that such a claim is meaningless within the Will Theory. But this derives directives with seemingly obnoxious binding force from a theory which has deprived itself of all possibilities of justification. And much the same is true, *mutatis mutandis*, for Teleological Theories. For a directive based on a certain type of practical argument could not have moral binding force and so could not be obligatory unless it were wrong not to follow directives arrived at by such a type of practical argument. The received theories distract attention from these questions by stressing the emotive appeal of such a method for generating obligation claims ("its rationality," "its autonomy," "its dignity") or the appeal to individuals of the particular obligations generated by such methods (that suicide is wrong, idleness is wrong, etc.).

How, then, can we avoid this epistemological circularity? Clearly, the procedure must be this. We must first show wherein the peculiar *binding force* implied in obligation claims consists. We must then demonstrate *the need for* claims with such a binding force. And lastly we must indicate *a method* for generating directives with such a binding force.

(IV—1) THE BINDING FORCE OF OBLIGATION CLAIMS. As argued earlier, the binding force of obligation claims is moral. We must now ask what this means. My answer is: "A directive has moral binding force" means, "It is not solely the addressee's business to decide whether or not to follow the directive." What, on my theory, gives directives moral binding force is not that they are sufficiently specific to permit us to establish them conclusively or unconditionally; not even that when they conflict with others, they are, or are rightly, regarded as overriding; it is, rather the fact that they concern themselves with issues and problems whose solution is *not solely the agent's business* but also that of others who have a legitimate concern about whether or not the person to whom such a directive applies follows it or not. When directives are of this sort it is justifiable for a society to take suitable measures to ensure that its members follow them. This seems to me the true kernel of the conviction, which runs through most versions of the Will

Theory, that the binding force of obligation claims lies in the fact that the addressees of such directives know that they are liable to incur the sanction if they do not follow them.

What is meant by the phrase, "not solely the agent's business to decide whether to follow the directive"? As children grow up, more and more of the things they have been directed to do at certain times and in certain ways are left to their own discretion. Eventually, whether they follow such directives becomes solely their business. At some stage, there comes an end to a mother's authority to see to it that her son, using the appropriate system of prudential and moral reasoning, keeps the right company, drinks the right drinks in the right quantities, and regularly attends the right church. All the same, we do not think there is an end to everyone's authority in regard to all directives. In regard to some, e.g., the law, it is not solely even an adult's business to decide whether or not to follow such directives.

That it is not solely his business to decide on some issue means that not all kinds of interference with his conduct in these matters are necessarily unjustifiable. Some may even be desirable. By "interference" I mean pressures which are in themselves obnoxious. I have in mind methods such as handcuffing, jailing, fining, and possibly simply "condemning," which require justification. I do not mean perfectly legitimate pressures such as reasoning or pleading with a person, which a mother may, perhaps should, exert even on an adult son.

However, the importance of ensuring compliance with certain directives may justify even interference. To show that some directive has moral binding force, it is not, therefore, enough to show that in a given society a person's failure to comply with such a certain directive is followed by group interference or that the group regards departure from such directives as a signal for hostility, to use Hart's phrase. If that is the case in a community, and if the community regards such hostility as justified, perhaps its absence as undesirable, then such a directive is indeed *regarded as* having moral binding force. And since such a practice tends to increase conformity, tends to act as a social pressure backing up the directive, we might say that, in a sense, such a directive *has* moral binding force. All the same, this answer which might be quite adequate in a sociological context, is not adequate in an epistemological one. The question we must answer is whether it is *rightly* so regarded. But are we not now

moving in a circle? Directives, we said, have moral binding force if the question of whether to follow them is not solely the agent's business. It is not solely the agent's business if it is permissible, perhaps desirable, for certain people to interfere for the purpose of ensuring conformity with the directive. But does not this mean simply that it *ought* to be someone's job or duty to see to it that addressees of such directives follow them? And does not this mean that those, whose job it is to see to this, *have an obligation* to do these things, which means in turn that they come under directives with moral binding force, which implies that someone else has an obligation to see that they follow these directives, and so on *ad infinitum*?

This looks like a vicious regress but it is not. Its appearance of viciousness may well be one of the reasons why some philosophers feel that "moral binding force" should be analyzed in terms of the *actual* social pressures operative in a community, or at most the community's actual *beliefs* of the appropriateness of such pressures, but never in terms of the *truth* of such beliefs. The realization that there is nothing vicious about this regress may help to make my own answer more acceptable. As I construe the claim "'Do *x*' has moral binding force," it implies that "'Do *x*'" is a directive in regard to which there *ought to be* a person whose job it is to ensure that all those to whom addressees of the directive applies follow it. But this means only that it is desirable that there should be such a person with such a job, if that were necessary to ensure that the directives are followed. And this may be true whether or not there is such a person. As the line of such supervisors lengthens, the likelihood increases that the remoter ones will carry out their duties of ensuring that those under them do so, without having themselves to be supervised by yet further ranks in the system. Thus, "'Do *x*' has a moral binding force" does not imply either that there is in fact a person who has such a job giving him the obligation to see to it that the appropriate persons do *x*, or that a certain person has an obligation to see to it that someone is appointed to such a job, let alone that everyone morally ought to take this job upon themselves. And so there is no vicious regress.

(IV—2) THE JUSTIFICATION OF MORAL DIRECTIVES. If my account of directives with moral binding force is correct, then they are regarded as other people's business, and so as licensing interference. Their existence thus narrows what a person is free to do, and so it is in need of justification.

Here is a sketch of such a justification. My argu-

ment takes a number of things for granted. It assumes, for instance, that a morality is an operative system of practical reasoning, that is, a system of general end-setting directives which the members of the group are taught in their childhood as part of the conventional wisdom of the group. Systems of practical reasoning differ from each other in respect to the way in which such directives can be supported. The best-understood system is that of self-interest. In that system, a general directive, such as "Protect your health," would be supported by showing that they are ways of attaining states of affairs (such as being able to do things one wants to do and to enjoy life) which are necessary conditions of the good life for the person in question. Such self-interested directives are capable of coming into conflict with a person's inclination. For a person to be able to follow such directives he must not only be taught what they are but must be trained in childhood to follow them even when they go counter to his inclinations. Such training will require certain forms of pedagogic interference during childhood. This we regard as justified on account of the great benefits which such training bestows on the individual, and because we know of no other way of enabling him to derive such benefits. Of course, since the benefits so derived come to *him*, the question of whether he follows such directives is solely his business, hence seeing to it that he later follows them is not justifiable. The main distinction to bear in mind is that between the proof of, or support for, the directives of self-interest, and the need (inherent in any form of practical reasoning) and the justification for interfering with a person's inclinations and desires. The first tells us *what* such directives are, and *why* they should be followed. The second tells us *the extent to which* they should be inculcated and enforced.

A system of general moral directives can be justified along similar lines, though we must bear in mind the differences as well as the similarities. The main similarities are first that the support for general moral directives presumably must lie in the fact that following them leads to certain desirable states of affairs, and secondly that, since they must be capable of coming into conflict with inclinations and desires, there is a need to train the young so that they know what such directives are, and are able to follow them when they conflict with inclination and desire. The main differences are the following: Moral directives must be regarded as capable of overriding not only inclination but also

the directives of self-interest when a person cannot follow both, hence it will be difficult to train the young to follow moral directives. This explains, though it does not by itself justify, the practice of embodying some of our most important moral directives (e.g., "Thou shalt not kill") in our legal system and supporting them by severe and supposedly effective sanctions. We do, however, believe in the justifiability of this practice of seeing to it that people follow moral directives even when they are grown up and have learned what is right and what is wrong. And we support this belief by saying that whether or not a person follows moral directives is not solely his business but other people's as well. And if this is true, then these others *are* entitled to see to it that people follow moral directives and, within limits, to take measures which are necessary to achieve this purpose.

We have brought to light an important difference between self-interest and moral directives. Both imply that the practice of teaching these directives, and using pedagogic molding techniques to make it possible for the young to follow them if they want to, can be justified by the improved lives of those involved. But there the similarity ends. For moral directives must be such as to be rightly regarded as overriding not only inclination but also self-interest; the question of whether or not someone follows them must not be solely his business; and so the social practice of seeing to it that everyone follows them must be justified (beyond the training period). Lastly, whereas we know quite well how we support or prove, and so how we formulate, the directives of self-interest, it is more difficult to do this for moral directives. Hence in the case of self-interested directives we can spell out why it is that they are rightly regarded as overriding inclination, and why society is justified in inculcating in the young an appropriate attitude toward these directives during their training but not later, and why it is that the matter of whether or not a person follows them is solely his business and no one else's, hence why no one has what are often misleadingly called "obligations to himself."²⁷ But in the case of moral directives it is not always easy to spell this out. What then is the formula for constructing moral directives? We already have the clues we need. We can derive it from the premiss that moral directives

must have a content such as to yield two things: (a) support for saying that they are rightly regarded as overriding inclination and self-interest, and (b) support for saying that whether or not someone follows them is not solely his business.

Let us try to derive such a formula. We can eliminate the solitary desert island, for in his case whether he follows these directives must be solely his business. In fact, the situation we must envisage is precisely that envisaged by Hobbes: a group of people following inclination except when self-interest conflicts with it, and having a will of the requisite quality to follow reason rather than inclination. Such people are then confronted by the question of whether these principles of action are adequate for the good life or not. Hobbes makes an excellent case for saying that they are not. It is based on the following sort of argument. (i) Human needs, wants, and aspirations can be better satisfied under conditions of proximity, specialization, and cooperation, than in isolation. (ii) However, the scarcity of goods and the resulting conflicts of interest, as well as the fear generated by the justified belief that others will follow aggressive inclinations, and the mutually conflicting directives of self-interest, tend to lead to harmful and wasteful expenditures of resources and an unbearable climate of life. (iii) If there were available for guidance a set of directives regarded as overriding the directives of self-interest and applicable on those occasions when following the latter would lead to such harmful and wasteful behavior, the climate of life and so life for everyone would be improved. (iv) However, if it is true, as is widely believed, that people tend not to do what they think is contrary to their interest, they will be tempted, even after training, not to follow the directives of morality but those of self-interest when the two conflict. (v) But since the behavior of a person who yields to the temptation to follow self-interest and to ignore moral directives will, *ipso facto*, detrimentally affect another person's interests, the question of whether or not he follows moral directives is *ipso facto* not solely his business but someone else's as well, namely, the business of the person whose interest would be adversely affected. And since such behavior, unless prevented, would adversely affect the climate of life, whether or not people follow moral directives, is everyone's business.

²⁷ Elsewhere I have dismissed this view much too cavalierly. There I was able to show merely that it is impossible to enter with oneself into the sort of temporary moral relationship into which one can enter with another. But this impossibility is perfectly compatible with one's having an *obligation* to do a certain thing simply because it is the best thing for one. On my present view, one cannot have such an obligation because it is no one else's business whether or not one does what is best for oneself.

(IV—3) THE MORAL "CALCULUS." In section (IV—2) we gave the outline of an argument showing the need, for a group of people following either inclination and/or a system of self-interested directives, to have a set of directives overriding self-interested ones, in regard to which the decision of whether or not to follow them is not solely the business of those to whom they apply. This outline justification shows that what creates the need for such an overriding system is the fact that following self-interested directives will often lead to harmful and wasteful conflict and, if this occurs generally, to a climate of life which fully deserves the title "cold war." These considerations yield up the missing item, the calculus we must employ to arrive at directives satisfying the criteria for being moral. Such a formula must spell out how we determine the content of the directive and the circumstances under which it applies.

It will be remembered that such directives must spell out what is to be done in circumstances when two people, following directives of self-interest, could not both attain their end and would be driven into mutually harmful or wasteful efforts to attain their own end while preventing the other from attaining his, and where such efforts would be undesirable, at any rate if prosecuted "with no holds barred." The content must indicate which one of the interests is to give way or what compromise is to be made.

To understand the peculiarity of moral rules, we must however bear in mind a second requirement. Where interests conflict, there are many possible regulations dealing with the conflict. The directive embodying the regulation would not be properly moral (as opposed to being legal or conventional) unless it *purported* to be the *best possible* way of regulating such a conflict. Hence, other things being equal, general moral directives are open to critical scrutiny and to modification as social conditions change and our knowledge of consequences increases. The rationale for this second requirement is, of course, that since moral directives from the nature of the case will override someone's concern every time they apply, everyone must have (as far as possible) the same good reason for accepting such a general regulation.

I offer a simple example: *A* and *B* are interested in renting apartments. Both have looked at a great many of them, but have liked a certain apartment best. *A* was the first to inspect it and to pay a

deposit. *B* saw it shortly afterwards, and to secure it, offers to pay a higher rent. The moral rule here is that the giving of a deposit *binds* the landlord and so excludes other prospective tenants. All concerned have the same good reasons for accepting this solution as always overriding. For although on this occasion, the rule excludes *B* and thus goes against *B*'s interest, it is still in *B*'s interest just as much as in *A*'s that there should *be* such a rule, which will protect him on future occasions when he is the first to find a suitable apartment. Hence allocating obligations is in the end justified by the benefit everyone receives from this practice.

In this essay I have advanced the following major theses:

(1) All obligation claims are subclasses of general directives with moral binding force, and so are an integral part of a morality even though some, e.g., promissory or legal obligations, assign tasks which, for being *thus* assigned, would not be moral tasks.

(2) The binding force which is characteristic of such moral directives can be characterized by saying that the question of whether to follow such directives or not is not solely the business of those to whom they apply.

(3) This characterization of the binding force of such moral directives provides the answers to three important ethical questions:

(a) Why is it desirable that certain general directives should be universally taught and stringently enforced? The answer is that on their being generally followed depends the general climate of life which is the springboard from which an individual can, in accordance with his abilities and tastes, build a worthwhile life for himself; and because in the absence of such enforcement the likelihood of their not being followed would be very great.

(b) What is the proper subject-matter for general moral directives? The answer is: the general and authoritative adjudication between conflicting types of individual interests and concerns.

(c) How can the content of such directives be correctly formulated? The over-all principles on which such formulae are constructed is that it should be the *best* solution, i.e., the one which provides for each of those whose concerns are affected, as far as possible, an equally good reason for accepting this adjudication between conflicting concerns.

III. ON THE PSYCHO-PHYSICAL IDENTITY THEORY

JAEGWON KIM

THIS paper aims at an interpretation and evaluation of the so-called Psycho-Physical Identity Theory of mind. In Part I, I examine one group of arguments often offered in support of the theory. These arguments share the characteristic of being based upon considerations of theoretical simplicity in science; roughly, they contend that the Identity Theory leads to a simpler and more fruitful structure of scientific theory than its rival theories. Thus, these arguments can be called "arguments from scientific simplicity." I dispute the cogency and strength of these arguments. In Part II, I raise some questions concerning the interpretation of the Identity Theory—in particular, questions concerning the notion of identity of events and states—and suggest some tentative answers. I then examine another type of argument offered in support of the theory to the effect that it leads to a simpler scheme of entities than its rival theories. An argument of this type can be called "an argument from ontological simplicity."

I

1. The Psycho-Physical Identity Theory asserts that the so-called mental states, such as feelings of pain and the having of an after image, are just states of the brain. Pain, for example, is taken to be just some not as yet completely understood state or process in the brain. Let us refer to this brain state allegedly identical with pain as "brain state *B*." The identity in question is explained as the "strict identity" of reference, and this notion is illustrated by examples such as the identity of the Morning Star and the Evening Star. Thus, the two expressions "pain" and "brain state *B*" are said to refer to or denote the same event or state, just as the expressions "the Morning Star" and "the Evening Star" refer to the same planet. Further, the pain-brain state *B* identity is said to be an empirical fact subject to factual confirmation and not something that can be ascertained *a priori*.

If pain is identical with brain state *B*, there must

be a concomitance between occurrences of pain and occurrences of brain state *B*—and presumably not between occurrences of pain in me and occurrences of brain state *B* in someone else, but between my pains and my brain states *B*. Thus, a necessary condition of the pain-brain state *B* identity is that the two expressions "being in pain" and "being in brain state *B*" have the same extension; namely, the following equivalence must hold: "For every *x*, *x* is in pain at time *t* if and only if *x* is in brain state *B* at time *t*." An equivalence statement of a similar sort will correspond to each particular psycho-physical identity statement. I shall refer to a statement of this kind as "a psycho-physical correlation statement."

It is clear that a psycho-physical correlation statement does not entail the corresponding identity statement—at least, the identity must be understood in such a way that it is not entailed by a mere correlation. For otherwise the Identity Theory would fail to be a significant thesis distinguishable from other theories of mind such as some forms of Interactionism and the Double-Aspect Theory. It is perhaps clearer that the identity entails the corresponding correlation, and at least to this extent, the identity statement has a factual component. Further, the correlation is the *only* factual component of the identity; the factual content of the identity statement is exhausted by the corresponding correlation statement.

It is often emphasized that a particular psycho-physical identity (e.g., pain and brain state *B*) is a factual identity. From this some philosophers seem to infer that the Identity Theory is an empirical theory refutable or confirmable by experience. This is misleading, however. To begin with, a particular psycho-physical identity statement is not confirmable or refutable *qua* identity statement; it is confirmable or refutable insofar as, and only insofar as, the corresponding correlation statement entailed by it is confirmable or refutable by observation and experiment. There is no conceivable observation that would confirm or refute the iden-

tity but not the associated correlation. Moreover, not only the psycho-physical identity statement, but also the corresponding "psycho-physical interaction statement," the corresponding "psycho-physical double-aspect statement," and so on, are all confirmable or refutable by fact. And the very same evidence will confirm all of them or none of them; the very same evidence will refute all of them or none of them. Thus, the pain-brain state *B* identity statement is not an empirical hypothesis vis-à-vis the corresponding correlation, interaction, and double-aspect statements.

An essentially similar comment is in order for the claim that the Identity Theory itself is an empirical theory. It is asserted¹ that the Identity Theory would be "empirically false" if there were mental states not associated with the brain, namely "disembodied" mental states. This is true, although how the existence of such states could be ascertained *empirically* is a mystery. However, what is often not noticed is that the existence of disembodied mental states would refute not only the Identity Theory but also the Double-Aspect Theory, Parallelism, Epiphenomenalism, and some forms of Interactionism; for it would contradict the general hypothesis of psycho-physical correlation, a fact assumed, and to be explained, by philosophical theories of mind and body. If there were no correlation at all between mental and physical events, there would be no need for a theory of mind-body relation. So, within the context of philosophical discussion, it is of no significance that the Identity Theory is a factually refutable theory: it is not an empirical hypothesis vis-à-vis its rival theories.

2. The proponents of the Identity Theory, however, will be quick to point out that the foregoing considerations issue from an excessively narrow conception of "factual support." They will probably concede that an identity statement has no more direct observational consequences than the corresponding correlation statement. But it may be that the inclusion of such statements within a scientific theory will effect significant simplification of the structure of the theory and lead to new laws and theories, new explanations and predictions. If these conjectures turn out to be true, it would be proper to claim a broad factual support

for the Identity Theory. Arguments of this kind have been offered by most adherents of the theory; even some critics of the theory have argued that certain developments and discoveries in science would increase the plausibility of the theory.

In "Minds and Machines,"² Hilary Putnam offers an argument of this nature. He cites two advantages for identifying the mental and the physical:

- (1) "It would be possible . . . to derive from physical theory the classical laws (or low-level generalizations) of common-sense 'mentalistic' psychology, such as: 'People tend to avoid things with which they have had painful experiences.'"
- (2) "It would be possible to predict the cases (and they are legion) in which common-sense 'mentalistic' psychology fails."³

Briefly, the argument is that we ought to identify—or, at least, we are permitted to identify—the mental with the physical to make possible the reduction of mentalistic psychology to some physical theory of the body, presumably neurophysiology. Such a reduction is claimed to have two benefits: to unify and simplify scientific theory, and to make new predictions possible. The benefits of theoretical reduction in science cannot be questioned; in particular, the reduction of mentalistic psychology to a physical theory of the body, if carried out, would be a major scientific achievement. The question, however, is whether or not such a reduction presupposes the identification of the mental with the physical.

The reduction of one scientific theory to another involves the derivation of the laws of the reduced theory from the laws of the theory to which it is reduced.⁴ If the reduction is to be genuinely inter-theoretic, the reduced theory will contain concepts not included in the vocabulary of the reducing theory, and these concepts will occur essentially in the laws of the reduced theory. Hence, if these laws are to be derived from the laws of the reducing theory in which those concepts do not occur, we shall need, as auxiliary premisses of derivation, certain statements in which concepts of both theories occur. We may refer to these statements as "con-

¹ See Jerome A. Shaffer, "Recent Work on the Mind-Body Problem," *American Philosophical Quarterly*, vol. 2 (1965), pp. 81-104, especially pp. 93-94.

² In S. Hook (ed.), *Dimensions of Mind* (New York, New York University Press, 1960).

³ *Ibid.*, pp. 170-171.

⁴ For an illuminating discussion of the problem of reduction in science, see Ernest Nagel, *The Structure of Science* (New York; Harcourt, Brace & World, Inc., 1961), chap. 11.

necting principles." Thus, the reduction of mentalistic psychology to neurophysiology will require connecting principles in which both mentalistic and neurophysiological concepts occur; they will enable us to move from neurophysiological premisses to mentalistic conclusions.

Putnam's claim, then, may plausibly be taken as asserting that psycho-physical identity statements can serve as such connecting principles, just as statements like "Gas is a collection of molecules" and "Temperature is the mean kinetic energy of molecules" serve as connecting principles in the reduction of classical thermodynamics to statistical mechanics. This is plausible enough, but it alone does not support the psycho-physical identification. What needs to be shown is that *unless* the identification is made, the derivation of mentalistic laws from neurophysiological laws is impossible. That is, it has to be shown that nothing less than psycho-physical identity statements will do as psycho-physical connecting principles. But it is dubious that this can be shown; in fact, psycho-physical correlation statements seem sufficiently strong to function as the requisite connecting principles.

Consider a simple example: the usual derivation of the Boyle-Charles law of the gas from certain statistical-mechanical assumptions about gas. Essential to this derivation is the assertion that the temperature of a body of gas is a constant times the mean translational kinetic energy of the molecules of the gas—that is, $(1/2)M\bar{v}^2 = (3/2)RT$. Now, in order to derive the Boyle-Charles law, it is sufficient to interpret this equation as asserting a mere correlation between the temperature and the mean kinetic energy of a gas, namely to the effect that whenever a gas has such-and-such temperature, it has such-and-such mean kinetic energy, and conversely. It is not necessary to interpret the equation to the effect that temperature *is* mean kinetic energy. The equation clearly does not assert this; it only asserts that the *value* of temperature is the same as the *value* of mean kinetic energy.

Similarly, it is plausible to suppose that, without identifying the mental with the physical, mentalistic psychology can be reduced to physical theory in the sense that given a suitable set of psycho-physical correlation statements, laws of mentalistic psychology can be derived from physical theory. If psycho-physical identity statements are sufficient

for such derivation, the corresponding psycho-physical correlation statements will do just as well. It is not easy to demonstrate this conclusively for the reason that it is not clear exactly what an identity statement asserts over and above the corresponding correlation statement. In Part II of this paper I shall claim that an identity statement involves the identification of properties; for example, the pain-brain state *B* identity involves the identification of the property of being in pain with the property of being in brain state *B*. On the other hand, I shall claim that the corresponding correlation statement involves only extensional identity of the two properties. If this construal is correct, it is evident that the correlation statement can do everything that the identity statement does on the further reasonable assumption that there are no "intensional" contexts in neurophysiology.

3. Herbert Feigl and J. J. C. Smart have offered a somewhat different reason for identifying the mental with the physical.⁵ They have argued that by such identification we are able to eliminate what they call "nomological danglers," irreducible and unexplainable psycho-physical laws. It is argued that the identification of pain with a brain process is justified by some kind of methodological principle of "parsimony" or "simplicity" in science. The reasoning behind this argument seems to be as follows.

A correlation statement cries out for an explanation: Why is it that whenever and wherever there is water, there is H_2O ? Why is it that whenever and only whenever a person has pain he is in some specific brain state? Now, according to this line of reasoning, we can answer these questions if, and perhaps only if, we accept the corresponding identity statements. That is, we shall answer: Because water *is* H_2O , because pain *is* brain state *B*, and so on. But how can we explain these facts of identity? The answer is that they are not in need of explanation, that they cannot be explained—not because we lack relevant factual or theoretical information, but because they are not the sort of thing that can be explained. It is nonsense to ask for an explanation of why Cicero *is* Tully, or why the Evening Star *is* the Morning Star; it is equally nonsensical to ask for an explanation of why water *is* H_2O , or why pain *is* brain state *B*. Water just *is* H_2O , and pain just *is* brain state *B*. Generally, most identity state-

⁵ H. Feigl, "The 'Mental' and the 'Physical'" in H. Feigl, M. Scriven, and G. Maxwell (eds.), *Minnesota Studies in the Philosophy of Science*, vol. 2 (Minneapolis, University of Minnesota Press, 1958); J. J. C. Smart, "Sensations and Brain Processes," *The Philosophical Review*, vol. 68 (1959), pp. 141-156; J. J. C. Smart, *Philosophy and Scientific Realism* (London, Routledge & Kegan Paul, 1963).

ments do not seem to be capable of functioning as the explananda of scientific explanations; and psycho-physical identity statements are not in need of any explanation at all. On the other hand, psycho-physical laws, not being identity statements, must either be explained by deduction from higher laws or be taken as fundamental, unexplainable laws of nature. And if they are to be deduced from higher laws, then at least some of these higher laws in turn must be psycho-physical statements, and so in any case we are left with fundamental and irreducible psycho-physical laws.

Thus, it turns out that by moving from correlation statements to identity statements we do not explain facts that were previously unexplained; rather, we make them "non-explainable." Now, the question is this: In what sense does this achieve scientific or theoretical simplicity of the sort desired in science? In what respect does it contribute to the unity and fruitfulness of the system of scientific laws and theories?

I think that the simplicity thus achieved is rather trivial and of minimal significance from a scientific point of view. To begin with, the explanation of a correlation by an identity—"Why is pain correlated with brain state *B*?" "Because pain is brain state *B*"—is trivial. The factual cash value of the identity is simply the correlation, and in terms of factual information we are simply repeating in the explanans what is supposed to be explained. This is a far cry from the usual kind of scientific explanation in which a fact or a regularity is explained by invoking more general and more comprehensive laws and theoretical principles far richer in factual implication and theoretical power than the explanandum. But further, the introduction of these identity statements does not produce simplicity in a theoretically meaningful sense. The essential import of reduction in science lies in that it achieves a more parsimonious set of primitive concepts and primitive assumptions. When optics is reduced to electromagnetic theory, we thereby reduce the number of independent factual commitments about the world; the reduction of thermodynamics to statistical mechanics yields the same kind of simplification. Previously we had two theories, each with its own postulates; now we have one.

But merely to replace correlation statements by identity statements does not effect this sort of simplicity. First, such replacement does not reduce the number of primitive concepts, for

mentalistic concepts remain nonsynonymous with physicalistic concepts. Second, it does not reduce the number of independent primitive assumptions, for factual identity statements simply replace the corresponding factual correlation statements. It yields neither economy of concepts nor economy of assumptions.

4. If the foregoing considerations are correct, why should we, it might be asked, accept such apparently noncontroversial identity statements as "Water is H_2O " and "Temperature is the mean kinetic energy of molecules"? Should we not in these cases, too, stop short of identification and be satisfied with correlation? I would claim that the water- H_2O identity is, indeed, disanalogous with the pain-brain state *B* case, and that the temperature-energy case is rather like the pain-brain state *B* case.

"Water" and " H_2O " (in the sense of "substance whose molecular structure is H_2O ") are both substantive expressions referring to physical things and not to properties, events, states, or the like. Any bit of water has a decomposition into H_2O molecules; the two occupy the same spatio-temporal volume. The reduction of macro-chemistry to micro-chemistry, which is in part based on such identities as that of water and H_2O , is an example of micro-reduction:⁶ the things in the domain of the reduced theory have a decomposition into proper parts that belong in the domain of the reducing theory. In this sense, water has a decomposition into H_2O molecules; gas a decomposition into molecules and atoms. The net effect of micro-reduction is the explanation of the properties of some entity on the basis of the properties of the parts of the entity. So, water is literally made up of H_2O molecules, and a body of gas, of molecules and atoms.

Temperature, however, is unlike water and gas. Temperature is not a thing that is made up of certain parts; we cannot pick out a bit of temperature or an instance of it and say that it is made up of mean kinetic energy. The domain of classical thermodynamics does not contain temperature in the way the domain of macro-chemistry contains water; rather, it contains gas, or bodies of gas, and temperature is a state variable whose values are used to characterize the thermodynamic states of a system—in other words, it is a property of the things in the domain. But it in itself is not a thing: it has no decomposition into mean kinetic energy.

Take pain: again, pain is not a thing. It is

⁶ See P. Oppenheim and H. Putnam, "Unity of Science as a Working Hypothesis" in H. Feigl, M. Scriven, and G. Maxwell, *op. cit.*

supposed to be an event or state; and we may take it as a property of living organisms. A pain has no parts—it has no decomposition into parts of the brain or into neurons. It only has a “participant,” the person (or the biological organism) who has the pain. This person has a decomposition into parts of his body, organs, tissues, cells, and so on. So, there is almost an exact analogy between the temperature-energy case and the pain-brain state case. A physical thing, such as a body of gas, has temperature, and temperature itself is not a thing. The physical thing having temperature has a decomposition into molecules, and these molecules collectively have a certain property, namely a certain value of mean kinetic energy. And there is a definite correlation between this property of the molecules and the property temperature of the physical thing. A biological organism, such as a man, has pain, and pain itself is not a thing. The biological organism having the pain has a decomposition into organs, tissues, and so on—and, in particular, into the brain and the nervous system as a whole. The brain and the nervous system have a certain property, say some patterns of electric pulses (“brain state *B*”), and there is a definite correlation between the two properties, the property of being in pain and the property of being in this kind of brain state.

Thus, on this view, micro-reduction is still possible, and the unification of the domains of various scientific disciplines is also possible by repeated micro-reduction of one discipline to another. What should be noticed here is that the micro-reduction of one theory to another does not require—nor does it sanction—the reduction of properties in the sense of identifying macro-properties with correlated micro-properties. I conclude, therefore, that the adherents of the Identity Theory can find no support in the considerations of simplicity or unity in the structure of scientific theory.

II

1. The Identity Theory asserts that pain is identical with brain state *B*. But what does this mean?

To say that pain is identical with brain state *B* is to make a general statement that each particular occurrence of pain is identical with some particular occurrence of brain state *B*, and also, conversely, that each particular occurrence of brain state *B* is

identical with some particular occurrence of pain. It is clearly not intended that Plato's pain is identical with Socrates' brain state *B*; but rather that Plato's pain is identical with his own simultaneous brain state *B*. Hence, to claim that pain is identical with brain state *B* is to claim, among other things, that the two statements “Plato is in pain (at time *t*)” and “Plato is in brain state *B* (at time *t*)” describe or refer to the same event or state.

But what are we to understand by this assertion that two statements describe the same event or state? Under what conditions do two singular statements—restricting ourselves to singular statements—describe or refer to the same event or state of affairs? An answer to this question will have the general form: “Statement *A* describes event *a* and statement *B* describes event *b*, and *a* is identical with *b*.” So two problems emerge: first, what particular event or state does a given singular statement describe or refer to, and second, under what conditions does the identity of events obtain?

To be told that event *a* and event *b* are the same event if and only if *a* and *b* share all properties in common gives us no real enlightenment; it gives us a definition, no doubt a valid one, but not a practically usable *criterion*, of the identity of events. I would like to see someone apply this definition to Plato's being in pain and Plato's being in brain state *B* and deliver an opinion as to their identity or non-identity. To say that two singular statements refer to or describe the same event if and only if they are logically equivalent is clearly inadequate for the purposes of the Identity Theory.⁷ For the identity of the mental and the physical is assumed to be a factual one and not a matter of logic or meaning.

I suggest the following procedure. First, what is an event or state? An event or state can be explained as a particular (substance) having a certain property, or more generally a certain number of particulars standing in a certain relation to one another. Suppressing reference to time, we may take the expressions of the following kind as designating-expressions for events and states: “*a*'s being *F*,” “*b*'s being *G*,” “*a* standing in relation *R* to *b*,” etc, where ‘*a*’ and ‘*b*’ refer to particulars and ‘*F*’, ‘*G*’ and ‘*R*’ to properties and relations. Thus, Socrates' being in pain, Socrates' being in brain state *B*, and Socrates speaking to Theaetetus are all events or states. Although we normally distinguish between events and states, or between

⁷ K. Popper appears to have this concept of event in *The Logic of Scientific Discovery* (New York, Basic Books, 1959), pp. 88–90. However, a precise interpretation of Popper is uncertain.

events, states, and processes, I shall not attempt such a distinction here; in discussing the mind-body problem, philosophers speak indifferently in terms of events, states, and processes, and the fate of the Identity Theory does not hinge on whether mental events or states are identified with physical events, states, or processes. It suffices if the Identity Theorist concedes, as I think he would, that among the things that he wants to identify are Socrates' being in pain and Socrates' being in brain state *B*. With this understanding let us hereafter speak in terms of events for the sake of brevity.

Under this conception of event, the following criterion of the identity of events naturally comes to mind: The event *a's being F* and the event *b's being G* are the same event if and only if either the statements "*a is F*" and "*b is G*" are logically equivalent, or else the particular *a* is identical with the particular *b* and the property of being *F* (*F*-ness) is identical with the property of being *G* (*G*-ness). The criterion can be generalized in obvious directions so as to cover "relational events" and "compound events"; but the simple special case is all we need for the purposes at hand. Thus, on this criterion, Cicero's being a bachelor is the same event (state) as Tully's being an unmarried adult male; the Morning Star emitting yellow light is the same event as Venus emitting light of the color of the sunflower.

A singular atomic statement involving a one-place predicate—again we need not consider more general cases—has the form "*a is F*," and we may say that the statement, if true, describes or refers to the event *a's being F*. It follows that two singular statements "*a is F*" and "*b is G*" describe or refer to the same event if the event *a's being F* and the event *b's being G* are the same. Or we may say: Two singular statements describe the same event if they assert truly of the same particular that the same property holds for it.

* The foregoing, which is a fragment of what is hoped to be a full systematic analysis, not here presented, of the concept of event, admittedly does not precisely coincide with the ordinary presystematic notion. (But then it is not clear that there is one, ordinary notion of event.) Some of the points at which my analysis deviates from it may be noted here. For example, Brutus' killing Caesar and Brutus' stabbing Caesar turn out, on the proposed criterion of event identity, to be different events, and similarly, "Brutus killed Caesar" and "Brutus stabbed Caesar" describe different events. Notice, however, that it is not at all absurd to say that Brutus' killing Caesar is *not the same as* Brutus' stabbing Caesar. Further, to explain Brutus' killing Caesar (why Brutus killed Caesar) is not the same as to explain Brutus' stabbing Caesar (why Brutus stabbed Caesar); also, to postdict one is not to postdict the other.

Such common notions as one description of an event being more detailed than another description of the *same* event, one description being more informative than another, and so on, have no immediate meaning under the proposed analysis. If these notions are to be clarified, a more comprehensive notion of event (say, "happening"), namely one in terms of which "Brutus killed Caesar" and "Brutus stabbed Caesar" can be said to be *about the same happening*, would have to be constructed, hopefully on the basis of the more atomistic concept of event used in this paper. Anyhow, the critical portions of the present paper do not depend on a full acceptance of the proposed analysis (see the end of the following section).

⁸ For example, Rudolf Carnap in *Meaning and Necessity* (Chicago, University of Chicago Press, 1947), pp. 16 ff.

In identifying a mental event with a physical event, the identity of the particulars involved in the events presumably is not at issue, unless one would want to say that the Socrates who has pain is different from the Socrates who is in brain state *B*. A radical Cartesian Dualist would claim that mental events necessarily occur to mental substance and physical events necessarily occur to material substance. Let us disregard this problem for the moment, however, and assume that both pain and brain state *B* can be attributed to the biological organism, Socrates. Then the problem of the identity of Socrates' being in pain and Socrates' being in brain state *B* reduces to the problem whether or not the property of being in pain and the property of being in brain state *B* are the same property.⁸

2. The problem of the identity of properties is a difficult one. Most writers⁹ take logical equivalence or cointensivity as the criterion of property identity. Under such a criterion, all property-identity statements would be either logically or necessarily true, if true, and logically or necessarily false, if false. This shows that logical equivalence or cointensivity is obviously too strong as a criterion of property identity for the Identity Theory.

On the other hand, if cointensivity is too strong, mere coextensivity is too weak. In asserting that pain is identical with brain state *B*, the Identity Theorist intends to assert more than that there is a concomitance between occurrences of pain and occurrences of brain state *B*. These considerations put the Identity Theorist in a quandary: In order to state his theory in an intelligible and nontrivial way, he must produce a criterion of property identity that is weaker than cointensivity but stronger than coextensivity. Can such a criterion be found?

The task seems difficult but perhaps not impossible. The Identity Theorist may take heart in the

fact that there are *prima facie* cases of nonanalytic and contingent property identity. The following are some of the representative examples:¹⁰

- (1) Blue is the color of the sky.
- (2) Black is the color of ravens.
- (3) The property designated by the English word "redness" is the same as the property designated by the German word "Rot."
- (4) Goodness is Plato's favorite property (i.e., the property Plato liked best).

I have tried to enumerate as many different kinds of factual property identity as I can think of. If we inspect these cases, one common characteristic is seen to emerge: in each case, at least one of the terms of the identity refers to a property via some particular(s) that stands in a certain definite relation to it. In the first two examples, properties are referred to on the basis of the particulars that *exemplify* them, as in "the color of the sky" and "the color of ravens." In the third, a property is referred to on the basis of a word that *designates* it. In the last example, reference is made to a property by way of an "intentional relation" in which a particular, Plato, stands to that property. Hence, a reasonable conjecture is that all contingent statements of property identity contain, essentially, some expression that refers to a particular or individual. This seems true, but I have no general argument to prove it. The converse of the conjecture seems more intuitively plausible: If a statement of property identity includes an essential reference to a particular, then it is nonanalytic and contingent.

These considerations are admittedly inconclusive; but perhaps it is not unwarranted to suppose that the identity of pain and brain state *B*, if there is such identity, is unlikely to turn out to be a contingent and nonanalytic identity of properties. Here, there is no mention of particulars in referring to the properties; nor any mention or use, implicit or explicit, of such relations as exemplification and designation, or of any intentional relation. At any rate, it seems evident that if the pain-brain state *B* identity is a case of factual property identity, it is unlike the usual examples of such identity and

would require an explanation and justification of a special nature. And if the Identity Theorist objects to our entire procedure leading to this problem of the factual identity of properties, he is invited to propose a more reasonable alternative analysis of the concepts of event and of event identity.

The analysis of event proposed above explains why some Identity Theorists¹¹ are anxious to eliminate mental properties or "features" as well as mental events and states. For to allow irreducible mental properties that are exemplified is to allow irreducible mental events and states. Indeed, the problem of the identity of properties seems to be one of the central problems that confront the adherents of the Identity Theory. Whether or not my analysis of event is generally acceptable, we can argue as follows: Suppose that the property of being in pain is not the same as the property of being in brain state *B*. Then, surely, Socrates' being in pain and Socrates' being in brain state *B* would have to count as distinct events. Presumably, the former is a mental event and the latter its correlated physical event, and the two are distinct. This contradicts the Identity Theory.¹²

3. The so-called location problem for mental events and states has perhaps been the strongest obstacle to the Identity Theory; the alleged difficulties raised by it seem to have persuaded more philosophers against the theory than any other single difficulty.¹³ As formulated by the critics of the theory, the objection runs as follows. If a mental state is to be identical with a physical state, the two must share all properties in common. But there is one property, spatial localizability, that is not so shared; that is, physical states and events are located in space, whereas mental events and states are not. Hence, mental events and states are different from physical ones. When it is retorted that some mental events like itches and some cases of pain have fairly determinate spatial locations, it is answered that a pain or an itch may be locatable but not *having a pain* or *being itchy*. An obvious rejoinder to this move is to point out that having a hand, weighing 145 pounds, having a temperature of 97 degrees, and other so-called physical states and events have no clear spatial

¹⁰ Some of the examples are adapted from N. L. Wilson, "The Trouble with Meanings," *Dialogue*, vol. 3 (1964), pp. 52-64.

¹¹ For example, see Smart, "Sensations and Brain Processes," *op. cit.*, pp. 148-150.

¹² J. A. Shaffer writes: "... we cannot avoid admitting at the least the existence of *nonphysical properties* or *features*, even if we give up nonphysical events as a different class from physical events" (Shaffer's italics); "Mental Events and the Brain," *The Journal of Philosophy*, vol. 60 (1963), p. 162. My claim is that we cannot admit nonphysical properties without admitting non-physical events.

¹³ See Norman Malcolm, "Scientific Materialism and the Identity Theory," *Dialogue*, vol. 3 (1964), pp. 115-125; J. A. Shaffer, "Recent Work on the Mind-Body Problem," *op. cit.*, pp. 96-98.

locations either. A hand can be located in space, but having a hand cannot; a brain can be located in space, but not a brain state; my body can be located in space, but not my body's weighing 145 pounds.

Thus, the inconclusiveness and weakness of this objection to the Identity Theory stems not so much from the possible locatability of mental states and events as from the vagueness of the general concept of spatial location for events and states. Of course, it must be admitted that we do locate explosions, fires, and deaths; but it takes only a moment's reflection to notice that we do not locate events as such. Rather, we locate events by locating the particulars or things that "undergo" them. Something explodes in an explosion, and the explosion is located where the thing that explodes is located; when there is a fire, something burns, and the fire is where the burning thing is; and, similarly, a death takes place where the dying man is located. Particulars are located first; events and states are located relatively to particulars. Or, we may say, particulars are the primary localizable entities; events and states are localizable only derivatively.¹⁴

In terms of our analysis of the concept of event and state, we may say that an event *a's being F* can be located derivatively at the place where the particular *a* is located. Then, what the critic of the Identity Theory who takes the location problem seriously must show is that a mental event *a's being M*, where *M* is some mental property, is non-spatial in that the particular *a* to which *M* is attributed is not a spatially localizable entity. Namely, in order to show that Socrates' being in pain is not spatially localizable, it must be shown, on this construal of the location of an event, that Socrates to whom the property of being in pain is attributed is not a spatially localizable entity. But in order to show this one must show or assume that the subjects of mental properties—or the subjects of mental events and states—are immaterial souls or mental substances in the full-fledged Cartesian sense.

The situation, therefore, seems to be this. Insofar as the notion of the location of an event is unclear and vague, it is not clear that all physical events and states have locations; and insofar as it is made clear—in terms of the location of particulars—the assertion that mental events and states lack spatial locations implies the Cartesian thesis of the

immaterial soul and unextended mental substance. Hence, the objection based on the location problem is unclear and therefore inconclusive, or it begs the question at issue.

4. Can the Identity Theory claim to involve a simpler, more parsimonious ontology than the Dualist Theories? Under that theory, there would be only one system of events, namely physical ones some of which are also mental events, rather than two distinct interacting, correlating, or paralleling systems of events. And there would be fewer events, too, pain and the corresponding brain state being counted as one. It must be granted, I think, that the scheme of entities countenanced by the Identity Theory is clearly simpler than, and at least as simple as, that to be assumed by any alternative theory. But exactly what sort of ontological economy is effected by the Identity Theory? Or, equivalently, what does "fewer events" mean? The analysis of event given earlier suggests an answer to this question.

We assume an ontological scheme that includes particulars (substances) and properties as basic entities or one that includes events in addition to particulars and properties. In either case, an event can be understood in the manner explained in earlier sections on the basis of particulars and properties; and the identity of events can be explained on the basis of the identity of particulars and of properties. Let *M* be some mental property and *P* some physical property, and let *a* and *b* be particular substances. Then, factual identification of the mental event *a's being M* and the physical event *b's being P* involves (1) the identification of the properties *M* and *P*, and (2) the identification of the particulars *a* and *b*. Accordingly, the identification of the two events results in the reduction of both particulars and properties.

However, the net amount of economy thus achieved will vary depending on the alternative theory of mind that is taken as the point of comparison. An opponent of the Identity Theory may be one of the following two kinds: (a) one who rejects both (1) and (2) above, and (b) one who rejects (1) but is willing to accept (2). Philosophers of the first kind can be called "Cartesians"; they deny not only that mental properties and physical properties are identical but also that the "subjects" of physical properties or events can be the "subjects" of mental properties or events. On this view, nothing that has some mental property can have a

¹⁴ This point is anticipated by P. F. Strawson. See his *Individuals* (London, Methuen, 1959), p. 57.

physical one, nor vice versa; unextended mental substances are the subjects of mental happenings and the extended, unthinking matter is the substratum of physical properties. Thus, the Cartesians represent the opposite extreme of the Identity Theory: their theory involves a bifurcated system of particulars and a bifurcated system of properties, and either bifurcation is sufficient to generate a bifurcated system of events.

But a radical Dualism of this form is not the only alternative to the Identity Theory. A sort of Dualistic Materialism results if one accepts the identity of the particulars involved in the two events but not the identity of the properties. A theory of this form is materialistic in that it allows only spatio-temporally localizable particulars; and it is dualistic in that mental events are countenanced as a distinct system of events from the system of physical events. Over such a theory, the net

simplicity of entities effected by the Identity Theory lies merely in the reduction of properties. Whether such an economy of entities is of much philosophical significance is a difficult question that cannot be settled here; perhaps, it cannot be settled at all. But we will do well to remind ourselves that the economy in question would have to be attained in the face of the extreme implausibility besetting the factual identification of mental properties with physical ones, and also, as I tried to show in Part I, that the economy has no scientific import and hence cannot be supported by scientific considerations. The slogan of ontological economy does not by itself sanction the identification of any two factually correlated properties. We clearly do not think that ontological economy of this kind would justify the identification of, say, thermal conductivity and electrical conductivity as one property on the basis of the Wiedemann-Franz law.

Brown University

IV. MOORE'S MODAL ARGUMENT

R. L. PURTILL*

IN his paper "Four Forms of Scepticism,"¹ G. E. Moore presents the following argument:

Russell says to me: "You don't know for certain that you heard the sound 'Russell' a little while ago, not even that there *was* such a sound, *because* in dreams we often remember things that never happened." In what way could the alleged reason, if true, be a reason for the conclusion? . . . Suppose that I have had experiences which resembled this one in the respect that I felt as if I remembered hearing a certain sound a little while before while yet it is not true that a little while before I did hear the sound in question. Does that prove that I don't know for certain now that I did hear the sound "Russell" just now? It seems to me that the idea that it does is a mere fallacy, resting partly on a confusion between two different uses of the words "possible" or "may." . . . The argument seems to me to be precisely on a par with the following: It is possible for a human being to be of the female sex; (but) I am a human being; *therefore* it is possible that I am of the female sex. The two premisses here are perfectly true, and yet obviously it does not follow from them that I do not know that I am not of the female sex. I do (in my view) happen to know this, in spite of the fact that the two premisses are both true; but whether I know it or not the two premisses certainly don't prove that I don't.

This is an argument which can be applied to a great number of sceptical arguments, all of which have the general form: "You have sometimes been mistaken about *X*, so it is possible that you are now mistaken about *X*." It is an argument which depends on an analysis of modal terms such as "possible" that occur in such sceptical arguments generally. I shall therefore call it "Moore's modal argument against scepticism," or for short "Moore's modal argument." This argument is at least partially a refutation by logical analogy: it is alleged that a sceptical argument attributed to Russell (hereafter called "the sceptic's argument") is precisely on a par with a case presented by Moore (hereafter called "the parallel argument")

which seems to be a clear case of an argument with true premisses and a false conclusion.

I wish to discuss the following questions:

- i. Is Moore correct in calling the parallel argument a fallacy, and is it a fallacy for the reasons given by Moore?
- ii. Is it true, as Moore alleges, that the sceptic's argument is "precisely on a par" with the parallel argument?
- iii. Can the analysis of the use of modal words such as "possible" in these arguments be extended to other philosophically interesting uses of such modal words?

I

Moore presents the following criticism of the argument which we have called "the parallel argument":

The conclusion *seems* to follow from the premisses because the premiss "It is possible for a human being to be of the female sex" or "Human beings may be of the female sex" is so easily falsely taken to be of the same form as "Human beings are mortal," i.e., to mean "In the case of every human being it is possible that the human being in question is of the female sex" or "Every human being *may* be of the female sex." If, and only if, it did mean this then the combination of this with the minor premiss "I am a human being" would the conclusion follow: It is possible that I am of the female sex; or I may be of the female sex. But in fact the premiss "Human beings may be of the female sex" does not mean "Every human being *may* be," but only "Some human beings *are*." "May" is being used in a totally different sense from any in which you could possibly assert of a particular human being "This human being *may* be so and so." And so soon as this is realized it is surely quite plain that from this, together with the premiss "I am a human being" there does not follow "I may be of the female sex." There may be something more in the argument . . .

* An earlier version of this paper was read at the Northwest Philosophy Conference at Reed College, Portland, Oregon, on April 24, 1965. I wish to thank the referee of the *American Philosophical Quarterly*, whose suggestions and criticisms enabled me to strengthen and improve the paper.

¹ First printed in *Philosophical Papers* (London, George Allen and Unwin, 1956), pp. 193-222.

than this simple fallacy. But I cannot see that there is anything more in it.

Now what is the "simple fallacy" which Moore says has been committed? It seems to me that Moore is thinking in terms of syllogistic arguments. He first suggests the valid argument:

Every human being (is a being which) may be of the female sex.

I am a human being.

(Therefore) I (am a being which) may be of the female sex.

He then alleges that the correct translation of the true statement "It is possible for a human being to be of the female sex" is not "Every human being (is a being which) may be of the female sex." But rather "Some human beings are (beings) of the female sex." The argument then becomes:

Some human beings are (beings) of the female sex.

I am a human being.

(Therefore) I (am a being which) may be of the female sex.

Since this argument commits the fallacy of four terms and that of undistributed middle, Moore is obviously justified in saying that the conclusion does not follow from the premisses. If Moore is right, then, the parallel argument is a glaring *non-sequitur*. Clearly the crucial point here is Moore's analysis of the premiss "It is possible for a human being to be of the female sex." Is there any analysis of this premiss, different from Moore's, which would save the validity of the parallel argument? I wish to consider two suggestions for such an alternative analysis.

The first suggestion I wish to consider is the following: By "It is possible for a human being to be of the female sex" is meant "For any human being, it is *logically* possible that that human being be of the female sex." Or in symbols:

$$(\forall x)(Hx \supset \Diamond Fx)$$

Where $Hx = x$ is a human being

$Fx = x$ is of the female sex

and $\Diamond p$ = it is logically possible that p .

This is surely true, for to deny it is to assert that there is some human being who cannot possibly be of the female sex, that is;

$$(\exists x)(Hy \ \& \ \sim \Diamond Fx)$$

which is the same as saying that there is some

human being who is necessarily not of the female sex,

$$(\exists x)(Hx \ \& \ \Box \sim Fx)$$

where $\Box p$ = it is logically necessary that p .

And surely the sex of any individual is always a matter of contingent fact and never a matter of logical necessity.

The parallel argument itself can now be symbolized (with m = Moore) as:

$$(x)(Hx \supset \Diamond Fx)$$

$$\frac{Hm}{\Diamond Fm}$$

The conclusion can be validly inferred from the premisses by universal instantiation and *modus ponens*. Thus, we have again a valid argument, and in fact the translation into symbolic form of Moore's valid syllogism. Since the premisses are true, the conclusion must necessarily be true. However, this is not at all an embarrassing consequence if by " $\Diamond Fm$ " we mean only that it is logically possible that Moore is of the female sex. For example, it is quite compatible with the assertion that Moore is in fact a male, and not of the female sex. In fact " $\sim Fm \ \& \ \Diamond Fm$ " would normally be understood simply as asserting that it was factually (and not logically) true that Moore was not of the female sex.

There might seem to be an air of paradox in asserting that Moore, a male, might be a female since this might be taken as implying that some male could be a female. But this is to confuse the legitimate inference of " $(\exists x)(\sim Fx \ \& \ \Diamond Fx)$ " from " $\sim Fm \ \& \ \Diamond Fm$ " (by existential generalization) and the illegitimate passage from " $\sim Fm \ \& \ \Diamond Fm$ " to " $\Diamond (\exists x)(\sim Fm \ \& \ Fm)$ " or " $(\exists x)\Diamond(\sim Fm \ \& \ Fm)$ " which being contravalid modal propositions cannot in fact be the conclusion of any sound inference.

The second suggestion for an alternative analysis is as follows: By "It is possible for a human being to be of the female sex" is meant "For any human being, there is a nonzero probability that that human being is of the female sex" ($(\forall x)[\text{Prob}(Fx/Hx) \neq 0]$). The evidence for this statement is that just mentioned by Moore: some human beings are in fact of the female sex. But the meaning of the statement "For any human being there is a nonzero probability that that human being is of the female sex," is not, contrary to what Moore says, merely that some human beings are of the female sex, but rather that *since* some human

beings are of the female sex, the conditional probability that something will be of the female sex on the evidence that it is a human being is not equal to zero.

Now if this interpretation of the crucial premiss is accepted, the parallel argument will not be a fallacy for the reasons given by Moore. But it will be erroneous for another reason. For this argument to be sound, it would have to be the case that *all* we knew about Moore was that he was a human being, while in fact we know a good deal more than this, and this further information affects the probabilities.

To clarify this point consider the following "cases":

Case 1 All we know about G. E. Moore is that G. E. Moore is a human being. In this case the probability that Moore is of the female sex is about .5, since about half of all humans are female.

Case 2 We know that G. E. Moore is a human being and that the initials G. E. stand for "George Edward." The probability that Moore is of the female sex is now quite small, but is not zero, since some females have male names.

Case 3 We know that G. E. Moore is the third son of D. Moore, M.D. The probability that Moore is of the female sex is now zero, since our information precludes this possibility.

Since we do have information that Moore is a male, this second suggestion for an alternative analysis fails to save the parallel argument. For on this interpretation of the premisses, the conclusion would follow from the premisses only if we lacked information which we in fact possess. Still the argument would be a valid one if it was interpreted as meaning: "The conditional probability that something will be of the female sex on the evidence that it is a human being is not equal to zero. Therefore *if* all that we knew of Moore were that he was a human being, *then* it would follow that there would be a nonzero probability that Moore was of the female sex." As with the previous alternative analysis we have here an argument which Moore might very well accept, but which is comparatively trivial. Thus neither alternative analysis of the crucial premiss produces a version of the parallel argument which is a good argument and at the same time leads to the conclusion that in

any interesting sense, Moore may be of the female sex.

II

Is the sceptic's argument in fact "precisely on a par with" the parallel argument? Moore states that it is, and gives a criticism of it on the same lines as his criticism of the parallel argument:

What really does follow from this premiss (that I have had experiences which resembled this one in the respect that I felt as if I remembered hearing the sound "Russell" a little while before, while yet it is not true that a little while before I did hear the sound "Russell") is this: That it is possible for an experience of a sort, of which my present experience is an example, i.e.,: one which resembles my present experience in a certain respect, *not* to have been preceded within a certain period by the sound "Russell." Whereas the conclusion which is alleged to follow is: It is possible that *this* experience was not preceded within that period by the sound "Russell." Now in the first of these sentences the meaning of "possible" is such that the whole sentence means merely: Some experiences of feeling as if one remembered a certain sound are not preceded by the sound in question. But in the conclusion: It is possible that this experience was not preceded by the word "Russell" or this experience *may* not have been preceded by the word "Russell"; "possible" and "may" are being used in an entirely different sense. Here the whole expression merely means the same as: "*It is not known* for certain that this experience was preceded by that sound." And how from "Some experiences of this kind were not preceded" can we possibly be justified in inferring "It is not known that this one was preceded"?

As before, if Moore is correct in saying that the crucial premiss means merely that "Some experiences of seeming to remember a certain sound are not preceded," then he is surely right in calling the sceptic's argument a fallacy. The valid argument would be:

All experiences of seeming to remember a certain sound are possibly not preceded by the sound in question.

This is an experience of seeming to remember a certain sound.

(Therefore) This is possibly not preceded by the sound in question.

And the invalid argument would be:

Some experiences of seeming to remember a certain sound are not preceded by the sound in question.

This is an experience of seeming to remember a certain sound.

(Therefore) This is possibly not preceded by the sound in question.

This again commits the fallacies of four terms and undistributed middle.

But let us consider again the two alternative analyses which failed to save the parallel argument. First, the crucial premiss in the sceptic's argument might be interpreted as meaning: "For any experience of seeming to remember hearing the sound 'Russell' it is *logically* possible that that experience was not preceded by the sound 'Russell'." Or in symbols:

$$(x) (Sx \supset \Diamond \sim Px)$$

Where $Sx=x$ is an experience of seeming to remember a certain sound and $Px=x$ is preceded by the sound in question

This premiss must be considered true unless we wish to assert its contradictory:

$$(\exists x) (Sx \& \sim \Diamond \sim Px)$$

which is the same as:

$$(\exists x) (Sx \& \Box Px)$$

That is, at least *some* experiences of this kind are necessarily preceded by the sound in question.

Would anyone wish to maintain this position? It might seem so. For someone might wish to claim that it is logically impossible that *all* experiences of seeming to remember a certain sound should not be preceded by the sound in question, and that therefore some experiences of seeming to remember a certain sound must necessarily be preceded by the sound in question. But whatever the virtues of this claim, it does not lead to the conclusion that $(\exists x) (Sx \& \Box Px)$.

The original claim, that it is impossible that all experiences of seeming to remember should not be preceded, can be symbolized:

$$\sim \Diamond (x) (Sx \supset \sim Px)$$

this is equivalent to

$$\Box \sim (x) (Sx \supset \sim Px)$$

which in turn is equivalent to

$$\Box (\exists x) \sim (Sx \supset \sim Px)$$

and finally to

$$\Box (\exists x) (Sx \& Px)$$

But this is neither equivalent to, nor does it imply

$$(\exists x) (Sx \& \Box Px)$$

Thus the assertion of

$$\sim \Diamond (x) (Sx \supset \sim Px)$$

gives no grounds for the assertion of

$$(\exists x) (Sx \& \Box Px),$$

and indeed it would seem that whether or not a given sound precedes an experience of seeming to remember is always a matter of fact and never a matter of *logical* necessity.

If we grant this, and allow the premiss

$$(x) (Sx \supset \Diamond \sim Px)$$

as true then the argument will proceed as in the case of the parallel argument. *St* will be true in specific instances, and in those cases the argument

$$(x) (Sx \supset \Diamond \sim Px)$$

St

$$\hline \Diamond \sim Pt$$

will be a valid argument with true premisses and prove the truth of the conclusion $\Diamond \sim Pt$.

But again, quite unembarrassingly; " $\Diamond \sim Pt$ " is quite compatible with "*Pt*" and indeed

$$Pt \& \Diamond \sim Pt$$

would be the standard way of saying that it was factually (and not logically) true that this experience was preceded by the sound in question. Nor need this have embarrassing consequences with regard to knowledge or certainty, since both knowledge and certainty are commonly claimed concerning factually true propositions.

Thus the sceptic's argument may very well be valid and have true premisses and therefore show that it is logically possible that an experience of seeming to remember was not preceded by the sound in question, without causing us any uneasiness about knowledge or certainty.

Now Moore might be willing to admit this, since in a number of his papers he seems to present, or be on the verge of presenting, an argument of this kind:

"Even if it is *logically* possible that a given statement *S*, is false, it is legitimate to claim that I am certain that *S*. For to deny the legitimacy of such a claim is either to confuse empirical certainty with *a priori* certainty of the sort found, e.g., in mathematics, or to make the dubious claim that mathematical certainty is the only 'true' certainty."

Whether or not Moore would advance such an argument it has been advanced by later philosophers and it seems to me that this ground has

been thoroughly gone over in recent philosophy,⁸ and so I would like to turn my attention to the second, and I think more interesting, interpretation of the major premiss in the sceptic's argument.

According to this interpretation, the major premiss of the sceptic's argument should be read as: "For any experience of seeming to remember that I heard the sound 'Russell' a little while ago there is a nonzero probability that that experience was not preceded within that period by the sound in question." ($(x) [\text{Prob} (\sim Px/Sx) \neq 0]$). In other words, the sceptic is alleging that merely on the evidence that I seem to remember hearing the sound "Russell" a little while ago, it is not certain that I did in fact hear the sound "Russell" within the period, or to put it another way the conditional probability that my experience was *not* preceded by the sound in question is not zero on the evidence given. This seems on the face of it to be a reasonable claim, and to be different from the claim that it is merely *logically* possible that the experience was not preceded by the sound in question. For to say that an event is logically possible seems to be only to say that the description of that event is not self-contradictory. But to say that an event has a nonzero probability on certain evidence is to say that you do not have conclusive empirical evidence to show that that event will not occur. This seems to be a stronger claim, a more interesting claim, and in fact may be what many sceptics have wished to claim about cases of the sort we are discussing.

At this point in the parallel argument, we showed that such an analysis of the major premiss of that argument would not save the argument, for we had information that did make the probability that Moore was of the female sex a probability of zero. Do we have any similar information in the case of the sceptic's argument?

It seems to me that Moore would want to say that we do have such information, but, that he fails to say what this information is, and by so failing, overlooks certain problems.

My reasons for saying that Moore would want to say that we do have such information are passages such as that quoted above in which Moore distinguishes between the experience which he is discussing and other experiences which resemble it in a certain respect. He says that because experiences

which resemble this experience in a certain respect were not preceded by the sound in question it does not follow that this experience was not so preceded. By this he seems to suggest that this experience may be different from the other experiences in some important respects. This would be exactly parallel to the argument that because Moore resembles other individuals in a certain respect, that of being a human being, and these individuals are of the female sex, it does not follow that Moore is of the female sex. We know further facts about Moore which show that he differs from these other individuals in important respects, and from this knowledge we infer that he is not of the female sex.

But what is the knowledge we have about the experience Moore is discussing which shows that it differs from the experiences that were not preceded by the sound in question? Moore fails to tell us, and by failing to tell us avoids the problems of specifying what such knowledge might be. Might it be that I remember "clearly and distinctly," or that my memory agrees with the remembrance of others? It seems very difficult to find any set of characteristics which might attach to such an experience and which would be such that it would make the probability that my experience of seeming to remember the sound in question was not preceded by that sound a probability of zero.

At this point Moore or the philosopher who, like Moore, wishes to reject scepticism, has two alternatives open to him. He may allege that there are some circumstances which, if known to be present, would make the probability that the experience of seeming to remember hearing a certain sound was not preceded by that sound a probability of zero. But if the philosopher takes this alternative, the sceptic will certainly argue that because we have been mistaken in the past in thinking such circumstances to be present when they were not, we cannot now be certain that such circumstances are present, and hence cannot be certain that the experience in question was preceded by the appropriate sound . . . and so on *ad infinitum*.

However the anti-sceptical philosopher might choose instead to argue as follows: "It is true that there is a probability greater than zero that I am mistaken in this instance. But there is an overwhelmingly greater probability that I am not mistaken. Since in the vast majority of cases when I

⁸ For example, Paul Edwards, "Bertrand Russell's Doubts About Induction" in *Logic and Language First Series*, ed. Antony Flew (Oxford, 1952), esp. p. 6E; R. B. Braithwaite, "Probability and Induction" in *British Philosophy at Mid-Century*, ed. C. A. Mace (New York, 1957), esp. pp. 146 ff.; Frederick L. Will, "Generalization and Evidence," *Philosophical Analysis* (Englewood Cliffs, New Jersey, 1963), esp. pp. 365 ff.

seem to remember hearing a certain sound, my experience of seeming to remember was preceded by the sound in question, I will presume that in this case it was so preceded. And since the probability that my experience was not so preceded is extremely small, I will have no qualms about saying that I am certain that my experience was so preceded or that I know that it was so preceded."

Such an answer, it seems to me, does full justice both to the sceptic's doubts and to Moore's very proper rejection of those doubts. The implications of such an answer are interesting ones, and I shall examine them briefly in my concluding section.

III

Sometimes it is argued³ that an answer to scepticism of the sort suggested above cannot be successful, for the sceptic can always reply to it by challenging the "principle of induction," upon which such probabilistic arguments are based. In its most minimal form the principle of induction can be stated as follows: "In certain circumstances, which can be specified, we are justified in reaching conclusions about unexamined cases on the basis of information about examined cases."

Hume's criticism of this principle is well known: it cannot be justified deductively and to attempt to justify it inductively is to argue in a circle. But it is not always sufficiently recognized that scepticism about the principle of induction is incompatible with the sort of sceptical argument which we have been considering. For the argument: "You have been mistaken about *X* in the past, so it is possible that you may be mistaken about *X* in the future," makes a conclusion about unexamined cases (that they may be mistakes) on the basis of examined cases (similar mistakes in the past). It is thus as dependent upon the principle of induction as any other inductive argument. So the sceptic who denies the principle of induction cannot use this argument.⁴

Notice the quite limited scope of this objection to the sceptic's procedure. It does not in itself question the sceptic's doubt of the principle of induction, it simply points out to the sceptic that he cannot "have it both ways." If he doubts the legitimacy of argument from examined cases to unexamined ones this applies just as much to arguments from past mistakes to the possibility of future mistakes as to

arguments from past successes to future successes. Any denial of the principle of induction, or suspension of judgment concerning it, casts doubt *equally* on all empirical propositions. Any argument from examined cases to unexamined is as bad, and therefore as good, as any other. Therefore the sceptic who denies the principle of induction or suspends judgment on it cannot without inconsistency use some inductive arguments (from past mistakes to the possibility of future ones) while at the same time rejecting inductive arguments in general, or at least suspending judgment on them.

This being so, there is a sense in which the sceptic who uses the argument from past mistakes is thereby committed to some form of the principle of induction. This does not mean, however, that he must accept the usual judgments of inductive arguments. A sort of intermediate scepticism is possible, which accepts arguments from examined cases to unexamined ones, but insists that extremely strict standards be applied to such arguments. Even within this intermediate variety of scepticism there are two important subspecies. One sort of intermediate sceptic, whom I shall call the Type 2 sceptic, will allow inference from examined to unexamined cases only where no error of any kind is possible, where by "possible" it seems fairly clear that he means "logically possible." The Type 2 sceptic is never satisfied by any demonstration that all practical possibility of error is ruled out, or that there is no reasonable ground (by ordinary standards) for suspecting error.

As opposed to the Type 1 sceptic, who denies the principle of induction or suspends judgment on it, the Type 2 sceptic is able to use the argument from past error, since his position allows him to use negative inductive evidence, in the following way. Suppose that the Type 2 sceptic wonders whether the fact that he seems to remember hearing the sound "Russell" a moment ago justifies the belief that the sound "Russell" occurred a moment ago. On his principles, he is not entitled to this belief if it is logically possible that he may be mistaken. The fact that mistakes have occurred in cases of this kind in the past shows that mistakes are logically possible in cases of this kind, and he concludes that he is not entitled to the belief in question. The past errors in other words serve as the grounds of an inference *ab esse ad posse*, from "*p*" to " $\Diamond p$." Thus the Type 2 sceptic can consistently use the argument from past

³ For instance, Bertrand Russell in Chap. VI of *Problems of Philosophy* (London, 1912), esp. pp. 68-69 (of the reset edition).

⁴ The classical case of the sceptic who denies the principle of induction but uses the argument from past errors is Hume himself. Cf. *An Enquiry Concerning Human Understanding*, Sects. IV and X.

error, as showing the logical possibility of error in the sort of case in question.

However, his position is open to the objection that he is confusing induction with deduction, or applying standards to inductive arguments which are applicable only to deductive ones. Since this argument is a rather familiar one,⁵ let me turn to the less frequently considered case of the other type of partial sceptic.

The Type 3 sceptic does not demand deductive certainty from inductive arguments. He admits that inductive arguments are by their nature such that there is a *logical* possibility of error. However, he insists that we reject, as untrustworthy, sources of information which in the past have led us into error.⁶ The argument from past error is, of course, the main weapon of the Type 3 sceptic. However, the Type 3 sceptic faces the following dilemma. He presumably wishes to argue from the fact of past error to the possibility of future error, and then to reject any source of information which carries this possibility of future error. However, if by "possibility" of future error he means only "logical possibility" his position becomes that of the Type 2 sceptic and is open to the objections to that position.

However, if by "possibility" of future error he means some nonzero *probability* of future error, he must face the following argument: Once we use the idea of probability we commit ourselves to certain consequences of this idea. Suppose that we have one case of error in one thousand cases of seeming to remember a sound. We cannot argue from this that the probability of error in future cases of this kind is nonzero without at the same time facing the fact that the probability of error is extremely small, and the fact that the probability that we shall not be mistaken in future cases of this kind is quite close to one.

The Type 3 sceptic, by using the idea of probability, has thereby committed himself to certain implications of that idea. Unlike the Type 2 sceptic, he cannot use inductive arguments only in a negative way to establish the logical possibility of error. The Type 3 sceptic wishes to assert more than this and wishes to assert that on the evidence of past errors we shall in fact commit errors in the

future in the sort of case in question. But the arguments, based on the idea of probability, which he uses to reach this conclusion also show that we shall in fact sometimes not commit errors in the future, and show that the relative frequency of error can be expected to be quite small. Now this does not make it impossible for the Type 3 sceptic to reject any source of information which has in the past led us into error, but it makes this move seem quite unreasonable.

Suppose for example that in a thousand cases of seeming to remember hearing a certain sound, the sound in question preceded the experience of seeming to remember in all but one case. If he argues from this that the probability of error is *not* zero, the sceptic seems committed to some view of what the probability is. (Otherwise his position becomes merely that of the Type 2 sceptic.) On any consistent view of what the probability of error is, it must be acknowledged that it is quite small. If it is quite small the probability of no error is quite large ($\text{Prob}(\bar{E}) = 1 - \text{Prob}(E)$). Thus the sceptic's position amounts in this case to saying "I will put *no* reliance on seeming to remember a sound, even though I admit that in almost all cases seeming to remember a sound will be preceded by the sound in question." This is not an impossible position but it is, I think, an unreasonable one. The sceptic seems so afraid of being mistaken that he will reject generally reliable sources of information just because they sometimes lead him into error.

Notice that this argument against the sceptic is quite independent of the argument given by other authors;⁷ that closer examination of individual cases of, e.g., seeming to remember a sound, may show that they have features not shared by those cases of seeming to remember a sound which were not preceded by the sound in question, features which account for our confidence that no error is committed in some cases. Rather, I am arguing that even if we could not locate the source of error and all the cases were seemingly indistinguishable, we would still be justified in trusting our memory of sounds if in almost all cases the sound in question did precede our experience of seeming to remember it.

⁵ Edwards (*op. cit.*) interprets Russell's position in *Problems of Philosophy* essentially as Type 2 scepticism.

⁶ This is, of course, the basic principle of Descartes' methodological scepticism. Cf. the *Meditations*, First Meditation, and the *Discourse on Method*, Pt. II.

⁷ Notably by Nicholas Rescher in "The Legitimacy of Doubt," *Review of Metaphysics*, vol. 13 (1959), pp. 226-234. But cf. also p. 443 of A. D. Woozley's edition of Thomas Reid's *Essays on the Intellectual Powers of Man* (London, 1941) where this line of argument is foreshadowed (in rather a different connection).

Finally, we may note that the point about probability can be applied *fortiter* to the Type 1 sceptic, the sceptic who denies the legitimacy of making inferences from examined cases to unexamined ones. Since the whole notion of applied probability depends on proceeding from examined cases to unexamined ones, the Type 1 sceptic can never use this notion of probability. Especially in questioning the principle of induction he cannot use "possibly" to mean "with a nonzero probability," for without the principle of induction such uses of "probability" are meaningless. Thus if the Type 1 sceptic says that the principle of induction may be mistaken or may lead us into error, or that "the future may not be like the past," he cannot without absurdity have any probabilistic analysis of "may" in mind.

We have now considered a number of things that might be meant by the sceptic when he says "It is possible that you may be mistaken." If he means by this, as Moore alleges, only that we have sometimes been mistaken in the past, then Moore shows that this alone cannot prove that we may be mistaken now. If the sceptic means only that it is logically possible that we are mistaken, then this

shows merely that we are considering empirical knowledge and not *a priori* knowledge. And I have been especially concerned to show that if the sceptic means to say that there is a nonzero probability that we are mistaken, then this argument can be turned against him: for there is a much greater probability that we are not, in the sort of cases we have considered.

Thus, we may issue the following challenge to the sceptic: "In none of the senses of 'possible' which we have considered is there any important sense in which the argument from past error shows that it is possible that we are *now* mistaken. So if you wish to say that there is some important sense in which the argument from past error shows that it is possible that we are now mistaken, the burden of showing what sense this can be rests on you. Failing a satisfactory account of this sense, you are open to the accusation of using the word 'possible' without a clear sense, that is, of talking nonsense."

Like Moore, I am willing to entertain the possibility that there may be more to the sceptic's argument than I have been able to discover. But, like Moore, I cannot see that there is any more to it.

V. ON THE LOGIC OF "INTRINSICALLY BETTER"

RODERICK M. CHISHOLM AND ERNEST SOSA

WE present, first, reasons for rejecting certain widely-held theses concerning the relations that hold among the concepts: intrinsically good, intrinsically bad, intrinsically indifferent, and intrinsically better. We then offer a sketch of a logical calculus designed to exhibit the relations that we do believe to hold among these concepts.

I. PHILOSOPHICAL NOTES

1. For purposes of illustration, we proceed from a hedonistic assumption. We assume that pleasure is intrinsically good and displeasure intrinsically bad; or, more exactly, that any state of affairs consisting of more pleasure than displeasure is intrinsically good, and that any state of affairs consisting of more displeasure than pleasure is intrinsically bad. And for simplicity of exposition, we shall also assume that pleasure is the *only* thing that is intrinsically good and displeasure the *only* thing that is intrinsically bad. What we shall say, however, may readily be accommodated to the assumption that there are other things—states of affairs other than those involving pleasure—that are intrinsically good, and other things that are intrinsically bad.

We are assuming, therefore, that there are states of affairs, some of which are exemplified and some of which are not exemplified. Ordinarily, one would not say of any unexemplified state of affairs (say, that of everyone being happy) that it is good, or bad, or better than some other state of affairs. One might say, instead, that the state of affairs is

one which *would* be good, or bad, or better than some state of affairs, if only it were exemplified. For convenience, however, we shall speak of unexemplified states of affairs as being good, or bad, or better than other states of affairs.

2. We also assume that the concepts of intrinsic goodness and intrinsic badness are clear—that the reader can understand what is meant by saying that pleasure as such, "considered by itself and as if alone," is good in itself or good as an end, and that displeasure as such, "considered by itself and as if alone," is bad in itself or bad as an end. Any possible world (given our hedonistic assumptions) is good just to the extent that it contains more pleasure than displeasure and bad just to the extent that it contains more displeasure than pleasure.¹ And we shall assume, finally, that if the concepts of intrinsic goodness and intrinsic badness are clear, then the concept of intrinsic betterness is also clear. For example, given our hedonistic assumptions, we may say that a state of affairs that involves pleasure and no displeasure is intrinsically better than one that involves neither pleasure nor displeasure; and one that involves neither pleasure nor displeasure is intrinsically better than one that involves displeasure and no pleasure.²

3. First, we shall attempt to indicate that certain principles about the logic of betterness are false if they are interpreted as principles about the logic of intrinsic betterness. We shall single out five principles in particular which are now widely held.³ They are:

¹ Compare G. E. Moore's explication of intrinsic goodness: "To say of anything, *A*, that it is 'intrinsically' good is equivalent to saying that, if any agent were a Creator before the existence of any world, whose power was so limited that the only alternatives in his power were those of (1) creating a world which consisted solely of *A* or (2) causing it to be the case that there should never be *any world at all*, then, if he knew for certain that this was the only choice open to him and knew exactly what *A* would be like, it would be his duty to choose alternative (1), provided only he was not convinced that it would be *wrong* for him to choose that alternative." *Moore* (1942).

² We thus use "better" more broadly than it is ordinarily used. In the interest of correct usage, we could replace "better" by the semi-technical term "preferable," and then take "*p* is better than *q*" to mean that *p* is good and preferable to *q*, and take "*p* is worse than *q*" to mean that *q* is bad and preferable to *p*.

³ Principles (1), (2), and (5) were suggested in *Schwarz* (1900) and in *Scheler* (1913–1914), and were explicitly affirmed in *Brogan* (1919). All five principles are affirmed in *Halldén* (1957) and all but principle (3) and the "if" half of (the biconditional) (4) in *Aqvist* (1963). *Von Wright* (1963) affirms analogous principles about preference. Finally, *Baylis* (1965) affirms (1). It should be emphasized, however, that most of these authors do not restrict themselves, as we do, to the concepts of *intrinsic* betterness or preferability. And one should expect that the logics of the four following concepts would be quite different: (i) *p* is intrinsically better than *q*; (ii) *p* is instrumentally better than *q*; (iii) *p* is preferred, as an end, by *S* to *q*; and (iv) *p* is preferred, as a means, by *S* to *q*.

- (1) p is good if and only if p is better than not- p .
- (2) p is bad if and only if not- p is better than p .
- (3) If p is better than q , then not- q is better than not- p .
- (4) p is better than q , if and only if the conjunction of p and not- q is better than the conjunction of q and not- p .
- (5) For every state of affairs p , either (i) p is good or (ii) p is bad or (iii) it is false that p is better than not- p and it is false that not- p is better than p .

(1) The first of the five principles we are rejecting may be abbreviated as follows:

$$Gp \equiv pP \sim p$$

(We use P for the relation of better than, or preferable to, and reserve B for badness.) The principle is clearly false, since there are states of affairs which are not themselves intrinsically good but which, nevertheless, are better than, or preferable to, their negations. One such state of affairs, given our hedonistic assumptions, is that of there being no unhappy egrets; this state of affairs is preferable to its negation, but it is not itself one that is intrinsically good. The latter point may be seen if we consider the fact that the state of there being no unhappy egrets is not itself one that would rate any possible universe a plus. To rate the universe a plus a state of affairs should involve pleasure or happiness and not merely the absence of displeasure or of unhappiness.⁴

- (2) The second principle, viz.,

$$Bp \equiv \sim pPp$$

should be rejected on similar grounds. That state of affairs consisting of there being no happy egrets (p) is one such that its negation is better than it; but it is not itself one that is intrinsically bad. To rate the universe a minus a state of affairs should involve displeasure or unhappiness, not merely the absence of pleasure or happiness.

- (3) The third principle is:

$$pPq \equiv \sim qP \sim p$$

This, too, is false; for, although that state of affairs consisting of there being happy egrets (p) is better than that one that consists of there being stones (q), that state of affairs that consists of there being no

stones ($\sim q$) is no better, or worse, than that state of affairs consisting of there being no happy egrets ($\sim p$).

- (4) The fourth principle is:

$$pPq \equiv (p \& \sim q) P (q \& \sim p)$$

This is false, since (a) there are substitution instances of this formula which are such that the expression to the right of the equivalence sign is true while the one to the left is false and (b) there are also instances such that the expression to the right of the equivalence sign is false while the one to the left is true. Thus (a) it is better that there be stones and happy egrets ($p \& \sim q$) than that there be no stones and no happy egrets ($q \& \sim p$); but there being stones (p) is no better, and no worse, than there being no happy egrets (q). And (b) it is better that Smith and Smith's wife be happy ($p \& q$) than that Smith be happy (p); but there is no good reason for supposing that that contradictory state of affairs, consisting of Smith and Smith's wife being happy while Smith is not happy ($p \& q \& \sim p$) is better than that state of affairs, consisting of Smith being happy and it being false that both he and his wife are happy [$p \& \sim (p \& q)$].⁵

- (5) The final principle is:

$$(p) [Gp \vee Bp \vee (\sim (pP \sim p) \& \sim (\sim pPp))]$$

And this is false, since there being no unhappy egrets is neither intrinsically good nor intrinsically bad and yet it is a state of affairs that is intrinsically better than its negation.

4. The fifth principle above is sometimes put by saying that every state of affairs is either good, bad, or indifferent, where "indifferent" is defined in the following way:

$$Ip \text{ FOR } \sim (pP \sim p) \& \sim (\sim pPp)$$

If we retain this definition of "indifferent," as we shall, we cannot say that every state of affairs is good, bad, or indifferent. But if we contrast indifference, with a closely related concept which we shall call "neutrality," we may say that every state of affairs is good, bad, or neutral. For a neutral state of affairs is simply one that is neither good nor bad, so that:

$$Np \text{ FOR } \sim Gp \& \sim Bp$$

(We shall propose an alternative formulation

⁴ Principle (1) is explicitly rejected in *Kraus* (1937), pp. 227, 392, and in *Katkov* (1937), p. 67. These two books are Volumes II and III, respectively, of the publications of the Brentano-Gesellschaft.

⁵ Hector Neri Castañeda makes a similar point in a review of *Halldén* (1957). (See *Castañeda* [1958], p. 266.)

below.) Every state of affairs that is indifferent (or "totally indifferent") is also one that is neutral, but some neutral states of affairs, as we have seen, are not indifferent.

5. Were we to adopt Bentham's misleading terms "negative goods" and "negative evils," we could find true, but misleading, interpretations of principles (1), (2), and (5) above. A "negative good," we could say, is a neutral state of affairs that has a bad negation; there being no unhappy egrets is thus a negative good. And a "negative evil," we could say, is a neutral state of affairs that has a good negation; there being no happy egrets would thus be a negative evil. We could then say, in behalf of principle (1), above, that p is better than not- p if and only if p is either a positive good (i.e., a good) or a negative good. We could say in behalf of principle (2), that not- p is better than p if and only if p is either a positive evil (i.e., an evil) or a negative evil. And we could say, in behalf of principle (5), that every state of affairs is either (i) a positive or a negative good, or (ii) a positive or a negative evil, or (iii) totally indifferent.

But these terms "negative good" and "negative evil" are misleading, for there is nothing intrinsically good about a negative good and nothing intrinsically evil about a negative evil. Universes, as we have said, deserve no merits because of their negative goods and no demerits because of their negative evils. And whereas what is good is better than what is bad, what is negatively good is no better than what is indifferent, and what is indifferent is no better than what is negatively bad.

6. The view that we have been criticizing, however, is attractive theoretically, in that it enables us to define "good" and "bad" in terms of "better"; " p is good," one could say, means simply that p is better than not- p , and " p is bad" means that not- p is better than p . In view of what we have been saying, is it still possible to define "good" and "bad" in terms of "better"? We suggest that it is. Let us say that a state of affairs is *good* provided it is better than some state of affairs that is indifferent, and let us say that a state of affairs is *bad* provided

that some state of affairs that is indifferent is better than it.⁶

Or, in symbols:

Gp FOR $(\exists q) (Iq \ \& \ pPq)$

Bp FOR $(\exists q) (Iq \ \& \ qPp)$

7. The view that we have criticized implies that every state of affairs falls into one or another of three possible categories—the good, the bad, and the indifferent. Our view implies that every state of affairs falls into one or another of seven possible categories, namely:

- (i) Gp and $B\sim p$
- (ii) Gp and $N\sim p$
- (iii) Np and $B\sim p$
- (iv) Np and $N\sim p$
- (v) Np and $G\sim p$
- (vi) Bp and $N\sim p$
- (vii) Bp and $G\sim p$

Presumably, if we are hedonists, we will hold that nothing falls into categories (i) or (vii), for no state of affairs involving more pleasure than displeasure will be such that its negation is a state of affairs involving more displeasure than pleasure.

Categories (iii) and (v), it might be noted, are those of "negative goods" and "negative evils," respectively.

8. We affirm the following five principles:

- (1) If a state of affairs p is better than a state of affairs q , then it is false that q is better than p .
- (2) If a state of affairs p is not better than a state of affairs q , and if q in turn is not better than a state of affairs r , then p is not better than r .
- (3) If a state of affairs p has the same value as its negation (i.e., if it is not better than its negation and if its negation is not better than it) and if a state of affairs q has the same value as its negation, then p has the same value as q .
- (4) If a state of affairs p is such that p is better than any state of affairs that is indifferent, then p is better than its negation.

⁶ Our definitions of these ethical concepts were suggested by the following definitions of the analogous psychological concepts: "Let us say that a man is *indifferent* toward a given state of affairs A , provided he does not prefer A to not- A and he does not prefer not- A to A . He may now be said to *favor* A provided that he prefers A to any state of affairs towards which he is indifferent; and he *opposes* A provided that any state of affairs towards which he is indifferent is one that he prefers to A " Chisholm (1964). These definitions of *favoring* and *opposing* (or, in one sense of the terms, of *desire* and *aversion*) are analogous to the definitions of *good* and *bad* except for using "any" where the latter would use "some." The reason for the difference is that, whereas the ethical relation connoted by "has the same value as" is transitive, the psychological relation connoted by "is not preferred by S to" is not. It is not possible for two things that are ethically indifferent to be such that one is better than the other, but it is possible for a man to be indifferent toward two things and yet prefer one to the other.

- (5) If a state of affairs p is such that any state of affairs that is indifferent is better than the negation of p , then p is better than its negation.

Taking these five principles as axioms, we are able to develop a logical calculus, which we now describe.

II. THE VALUE CALCULUS (VC)

1. Elements

The elements of a pure system of extended propositional calculus (EPC)⁷, plus: ' P ', a two-place predicate constant meaning "is intrinsically preferable to."

2. Formation Rules

If C and D are wffs of EPC, then CPD is a wff. If L and M are wffs (of VC), then so are $\sim L$, $L \vee M$, and $(\exists p)M$.

('C' and 'D' are metalinguistic variables ranging over formulas of EPC. The logical constants here function autonomously. 'L' and 'M' are metalinguistic variables ranging over formulas of VC.)

3. Definitions

The usual definitions of ' $(p) \dots p \dots$ ' in terms of ' \sim ' and ' $(\exists p) \dots p \dots$ '; and of '&', ' \supset ', and ' \equiv ' in terms of ' \sim ' and ' \vee '; plus the following five definitions:

- (D₁) pSq FOR $\sim(pPq) \ \& \ \sim(qPp)$
 (D₂) Ip FOR $\sim(pP\sim p) \ \& \ \sim(\sim pPp)$
 (D₃) Np FOR $(\exists q)(Iq \ \& \ pSq)$
 (D₄) Gp FOR $(\exists q)(Iq \ \& \ pPq)$
 (D₅) Bp FOR $(\exists q)(Iq \ \& \ qPp)$

The letters 'S', 'I', 'N', 'G', and 'B' are to be read, respectively, as "is the same in intrinsic value as," "is intrinsically indifferent in value," "is intrin-

sically neutral in value," "is intrinsically good," and "is intrinsically bad."

4. Axioms⁸

- (A₁) $(p)(q)[pPq \supset \sim(qPp)]$
 (A₂) $(p)(q)(r)[(\sim(pPq) \ \& \ \sim(qPr)) \supset \sim(pPr)]$
 (A₃) $(p)(q)\{[\sim(pP\sim p) \ \& \ \sim(\sim pPp) \ \& \ \sim(qP\sim q) \ \& \ \sim(\sim qPq)] \supset [\sim(pPq) \ \& \ \sim(qPp)]\}$
 (A₄) $(p)\{(q)[(\sim(qP\sim q) \ \& \ \sim(\sim qPq)) \supset pPq] \supset pP\sim p\}$
 (A₅) $(p)\{(q)[(\sim(qP\sim q) \ \& \ \sim(\sim qPq)) \supset qP\sim p] \supset pP\sim p\}$

5. Rules of Inference

- (R₁) If for each of the propositional variables in a theorem of EPC we substitute a wff of VC, the universal closure of the resulting formula, with respect to each of its variables, is a theorem of VC.
 (R₂) If for any of the variables in any of the theorems of VC we substitute a wff of EPC, the universal closure of the resulting formula, with respect to each of its variables introduced via the substitution, is a theorem of VC.
 (R₃) *Modus ponens*.
 (R₄) If C is a theorem of VC; if $D \equiv E$ is a theorem of EPC; if F is a wff of VC; and if F differs from C only in having E in one or more places where C has D and in having a universal quantifier at its beginning for any variables in E which are not in C ; then F is a theorem of VC.

6. Theorems

- (T₁) $(p)[pS\sim p \equiv Ip]$ (D₁), (D₂)
 (T₂) $(p)(q)[(Ip \ \& \ Iq) \supset pSq]$ (A₂), (D₁), (D₂)
 (T₃) $(p)\{(q)\{[Iq \supset pPq] \vee (q)[Iq \supset qP\sim p]\} \supset pP\sim p\}$ (A₄), (A₅), (D₂)

⁷ See Church (1956), pp. 151-154. Strictly speaking what we use here is a combination of systems. On the one hand we have a standard system of extended propositional calculus (propositional calculus with quantifiers for propositional variables). On the other, a reinterpreted system of extended propositional calculus, such that its variables range not over propositions, but over states of affairs and such that where the propositional calculus had truth-functional operators it has existence or exemplification functional operators. Since the ambiguities involved are not liable to mislead, we have thought to save the reader the tedium of having all the distinctions drawn out in detail. Roughly, however, here is the crux of the matter. It is surely not propositions that are better or worse than one another, but states of affairs. The variables flanking occurrences of "P" must therefore range over the latter and not the former. Yet in our value calculus we want to make assertions: to state axioms and derive theorems. And in doing so we want, of course, to use logical machinery already developed; in particular a system of extended propositional calculus. As a consequence we have a systematic ambiguity involving our logical constants. Consider, for example: $[(p \vee \sim q) Pr] \vee \sim[spP]$. In this formula ' \vee ' and ' \sim ' in their first occurrences function as exemplification functional operators, operating on states of affairs to yield compound states of affairs; but in their second occurrence, they function as truth-functional operators, operating on propositions to yield compound propositions.

⁸ By means of our definitions axioms (A₃) — (A₅) can be shortened so that (A₃) reads like (T₂), and the conjunction of (A₄) and (A₅) reads like (T₃). These theorems, (T₂) and (T₃), may make the intuitions behind (A₃) — (A₅) more readily apparent.

- (T₄) $(p)(q)(r)[(pSq \& qSr) \supset pSr]$ (A₂), (D₁)
 (T₅) $(p)(q)(r)[(pPq \& qPr) \supset pPr]$ (A₂), (A₁)
 (T₆) $(p) \sim (pPp)$ (A₁)
 (T₇) $(p)pSp$ (T₆), (D₁)
 (T₈) $(\exists p)Ip$ (T₃), (A₁)
 (T₉) $(p)(q)[pSq \equiv qSp]$ (D₁)
 (T₁₀) $(p)(q)(r)[(pPq \& qSr) \supset pPr]$ (See Appendix)
 (T₁₁) $(p)(q)(r)[(pPq \& rSp) \supset rPq]$

Proof:

Similar to that for (T₁₀)—given in the Appendix—except for use of (T₄) instead of (T₈) in step (vi).

- (T₁₂) $(\exists q)[Iq \& pPq] \equiv (q)[Iq \supset pPq]$
 (T₁₃) $(\exists q)[Iq \& pSq] \equiv (q)[Iq \supset pSq]$
 (T₁₄) $(\exists q)[Iq \& qPp] \equiv (q)[Iq \supset qPp]$

Note: The proofs for (T₁₂)–(T₁₄) are already given as a pair of arguments by the Rule of Deduction, assuming one side of the equivalence as hypothesis and deducing the other side from it.

- (T₁₅) $(p)[(Gp \vee B \sim p) \supset pP \sim p]$ (T₃), (T₁₃), (T₁₄), (D₄), (D₅)
 (T₁₆) $(p)[Ip \supset Np]$ (T₂), (T₁₃), (D₃)
 (T₁₇) $(p)[(Np \& Nq) \supset pSq]$ (D₃), (T₂), (T₄), (T₆)
 (T₁₈) $(p)[(Np \& N \sim p) \equiv Ip]$ (T₁₆), (T₉), (T₁₇)
 (T₁₉) $(p)(q)[pPq \vee qPp \vee pSq]$ (D₁)
 (T₂₀) $(p)[Gp \vee Bp \vee Np]$ (T₈), (T₁₉), (D₃), (D₅)
 (T₂₁) $(p)[pP \sim p \supset (Gp \vee B \sim p)]$ (See Appendix)
 (T₂₂) $(p)[\sim(Gp \& G \sim p)]$ (T₁₅), (A₁)
 (T₂₃) $(p)[\sim(Bp \& B \sim p)]$ (T₁₅), (A₁)
 (T₂₄) $(p)[\sim(Gp \& Bp)]$ (T₁₅), (A₁)

It may be recalled that in formulating our (expository) hedonistic assumption, we said: "any state of affairs consisting of more pleasure than displeasure is intrinsically good and any state of affairs consisting of more displeasure than pleasure is intrinsically bad." Had we said instead "any state of affairs involving pleasure is intrinsically good and any state of affairs involving displeasure is intrinsically bad," then there would be ground for questioning (T₂₄); for in such a case that complex state of affairs consisting of Jones being happy and Smith being sad might conceivably be said to

be both intrinsically good and intrinsically bad. Similar observations apply to (T₂₅) and (T₂₆).

- (T₂₅) $(p)[\sim(Gp \& Np)]$ (D₃), (D₆), (T₁₃), (D₁)
 (T₂₆) $(p)[\sim(Bp \& Np)]$ (D₃), (D₆), (T₁₄), (D₁)

With this theorem—and T₂₀, T₂₄, and T₂₅—we see that the concept of the neutral, defined in our calculus (i.e., the neutral as being that which has the same value as something that is indifferent), is the same as the concept of the neutral that was defined in the first part of the paper (i.e., the neutral as being that which is neither good nor bad). Theorem 20, according to which every state of affairs is either good, bad, or neutral, should not be confused with the thesis, rejected at the outset, according to which every state of affairs is either good, bad, or indifferent.

- (T₂₇) $(p)[(Gp \& \sim Gq) \supset pPq]$ (D₄), (T₁₃), (D₁), (T₅), (T₁₀)
 (T₂₈) $(p)(q)[(Gp \& G \sim q) \supset pPq]$ (T₂₃), (T₂₇)
 (T₂₉) $(p)(q)[(Bp \& \sim Bq) \supset qPp]$ (D₄), (T₁₄), (D₁), (T₅), (T₁₀)
 (T₃₀) $(p)(q)[(Bp \& B \sim q) \supset qPp]$ (T₂₃), (T₂₉)
 (T₃₁) $(p)(q)[(Gp \& qPp) \supset Gq]$ (T₂₇), (A₁)
 (T₃₂) $(p)(q)[(Bp \& pPq) \supset Bq]$ (T₂₉), (A₁)
 (T₃₃) $(p)(q)[(Np \& Iq) \supset pSq]$ (D₃), (T₁₃)

With this theorem we justify our assertion that so-called "negative goods," neutral things with bad negations, are no better than what is indifferent, and that so-called "negative evils," neutral things with good negations, are no worse than what is indifferent.

- (T₃₄) $(p)(q)[(Gp \& Iq) \supset pPq]$ (D₄), (T₁₃)
 (T₃₅) $(p)(q)[(Bp \& Iq) \supset qPp]$ (D₅), (T₁₄)
 (T₃₆) $(p)(q)[(Gp \& Bq) \supset pPq]$ (T₂₄), (T₂₇)
 (T₃₇) $(p)(q)[(Gp \& Nq) \supset pPq]$ (T₂₅), (T₂₇)
 (T₃₈) $(p)(q)[(Bp \& Nq) \supset qPp]$ (T₂₆), (T₂₉)
 (T₃₉) $(p)(q)[(Np \& qPp) \supset Gq]$ (T₈), (T₃₃), (T₁₀), (D₄)
 (T₄₀) $(p)(q)[(Np \& pPq) \supset Bq]$ (T₈), (T₃₃), (T₁₁), (D₅)
 (T₄₁) $(p)(q)[Nq \supset (Gp \equiv pPq)]$ (T₃₇), (T₃₉)
 (T₄₂) $(p)(q)[Nq \supset (Bp \equiv qPp)]$ (T₃₈), (T₄₀)

Twenty-eight other theorems can be obtained by application of contraposition to (T₂₇)–(T₄₀). None of these is counterintuitive.

We prove, finally, that with respect to intrinsic value, every state of affairs falls into one or another of seven possible categories:

$$(T_{43}) (p)[(Gp \& B \sim p) \vee (Gp \& N \sim p) \vee (Np \& B \sim p) \vee (Np \& N \sim p) \vee (Np \& G \sim p) \vee (Bp \& N \sim p) \vee (Bp \& G \sim p)]^9 \quad (T_{20}), (T_{22}), (T_{23})$$

Brown University

APPENDIX

We give sample proofs for two of our theorems. (In what follows we omit universal quantifiers occurring before a complete line of formula, with some exceptions in the interest of clarity.)

$$(T_{10}) (p)(q)(r)[(pPq \& qSr) \supset pPr]$$

(i) $(pSr \& rSq) \supset pSq$	(T_4)
(ii) $(pSr \& qSr) \supset pSq$	(i), (T_9)
(iii) $(pSr \& qSr) \supset \sim(pPq)$	(ii), (D_1)
(iv) $(pPq \& qSr) \supset \sim(pSr)$	(iii)
(v) $(pPq \& qSr) \supset (pPr \vee rPp)$	(iv), (D_1)
(vi) $(rPp \& pPq) \supset rPq$	(T_5)
(vii) $rPp \supset \sim(rSq)$	(D_1)
(viii) $(rPp \& pPq) \supset \sim(rSq)$	(vi), (vii)
(ix) $(rSq \& pPq) \supset \sim(rPp)$	(viii)
(x) $(pPq \& qSr) \supset \sim(rPp)$	(ix), (T_9)
(xi) $(pPq \& qSr) \supset pPr$	(v), (x)

$$(T_{21}) (p)[pP \sim p \supset (Gp \vee B \sim p)]$$

(i) $pP \sim p \supset (\sim Np \vee \sim N \sim p)$	(T_{18})
(ii) $(pP \sim p \& \sim N \sim p) \supset (\sim(\sim pPp) \& \sim N \sim p)$	(A_1)
(iii) $\sim(\sim pPp) \supset \sim G \sim p$	(T_{15})
(iv) $(pP \sim p \& \sim N \sim p) \supset (\sim G \sim p \& \sim N \sim p)$	(ii), (iii)
(v) $(pP \sim p \& \sim N \sim p) \supset B \sim p$	(iv), (T_{20})
(vi) $(pP \sim p \& \sim Np) \supset (\sim(\sim pPp) \& \sim Np)$	(A_1)
(vii) $\sim(\sim pPp) \supset \sim Bp$	(T_{18})
(viii) $(pP \sim p \& \sim Np) \supset (\sim Bp \& \sim Np)$	(vi), (vii)
(ix) $(pP \sim p \& \sim Np) \supset Gp$	(viii), (T_{20})
(x) $pP \sim p \supset [(\sim N \sim p \supset B \sim p) \& (\sim Np \supset Gp)]$	(v), (ix)
(xi) $pP \sim p \supset (Gp \vee B \sim p)$	(i), (x)

BIBLIOGRAPHY

- AQVIST, L. "Deontic Logic Based on a Logic of 'Better'," *Acta Philosophica Fennica*, fasc. xvi (Helsinki, 1963).
- BAYLIS, Charles "Tranquility Is Not Enough," *Pacific Philosophy Forum*, vol. 3 (1965).
- BROGAN, A. P. "The Fundamental Value Universal," *Journal of Philosophy, Psychology and Scientific Methods*, vol. 16 (1919).
- CASTAÑEDA, Hector N. "Review of 'On the Logic of Better'," *Philosophy and Phenomenological Research*, vol. 19 (1958).
- CHISHOLM, R. M. "The Descriptive Element in the Concept of Action," *The Journal of Philosophy*, vol. 61 (1964).
- CHURCH, Alonzo *Introduction to Mathematical Logic* (Princeton, 1956).
- HALLDÉN, S. *On the Logic of 'Better'* (Uppsala, 1957).
- KATKOV, Georg *Untersuchungen zur Werttheorie und Theodizee* (Brünn, 1937).
- KRAUS, Oskar *Die Werttheorien* (Brünn, 1937).
- MOORE, G. E. "A Reply to My Critics" in *The Philosophy of G. E. Moore*, P. A. Schilpp (ed.) (Evanston, 1942).
- SCHELER, Max *Der Formalismus in der Ethik und die materiale Wertethik* (Halle, 1913-1914).
- SCHWARZ, Hermann *Psychologie des Willens zur Grundlegung der Ethik* (Leipzig, 1900).
- VON WRIGHT, G. H. *The Logic of Preference* (Edinburgh, 1963).

* It seems possible to embed our axiom system in a first order predicate calculus: Let the individual variables range over states of affairs. Let there be a unary function letter (abbreviated as $\sim x$) and a binary function letter (abbreviated as $x \vee y$), defined over the individual variables. Finally, let there also be a binary predicate letter (abbreviated as xPy). Our five definitions and our five axioms can then be stated within a first order predicate calculus. (We are indebted to L. Jonathan Cohen, A. Garfinkel, P. T. Geach, Francis W. Irwin, Jaegwon Kim, R. Duncan-Luce, and A. N. Prior.)

VI. OTHER MINDS AND THE USES OF LANGUAGE

DWIGHT VAN DE VATE, JR.

LANGUAGE, Wittgenstein held, has no essence. He saw language as a multifunctional instrument whose uses may be catalogued and separately clarified, but not ranked in a hierarchy nor reduced to one another. The word "language," like the word "game," is "open-ended." There are card games, football games, war games, party games, and so on. The many kinds of games have no single feature in common, but resemble one another as the members of a family might, being alike now in this feature, now in that feature, now in another. We cannot say in advance what we shall call a "game," nor can we lay down *a priori* what language must be. Language is, simply, what we *do*—with words written and spoken, gestures, facial expressions, and other sorts of signs.¹

Language, however, is what distinguishes human from animal behavior. Language-games are forms of human life. The idea that man has an essence is an old one, and if it is a reasonable one, it would seem equally reasonable to expect human features to be essential to a human activity.

My purpose in this paper is to show the sense in which language-games have a single final cause, that is, the sense in which language has an essence. The fact that language is a human activity entails that a single (normally covert) purpose underlies its many explicit uses and functions. I call this purpose "personification." Language is dramatic and social: one speaks (writes, gestures, etc.) in order to characterize oneself, in order to create an image of oneself and one's social context in one's own mind and in the minds of others. Every utterance is intended to personify, whatever other functions it may have. I approach this thesis in a roundabout way. In Section I, I argue that Wittgenstein's well-known refutation of solipsism is circular. Then in Section II, I refute my own refutation of Wittgenstein's refutation, drawing the moral that the intent of language is to personify.

I

Wittgenstein's attack on solipsism in the *Philosophical Investigations* is indirect. The solipsist contends that his experience might be insurmountably private, so that therefore he can never be certain of the existence of other persons. Wittgenstein argues that this contention cannot meaningfully be stated. To state it meaningfully, the solipsist must suppose that the language in which he states it is (or could be) a *private* language, a language insusceptible of comprehension by persons other than himself. The solipsist thinks of sense-data as accessible only to first-person observation, and he thinks of the vocabulary of his language as constructed by private ostensive definitions of sensation-words. Thus he pictures himself constructing a language by attaching arbitrary labels to private feelings. Wittgenstein thinks that this is a misleading picture of language.

It is misleading because language is a rule-directed activity, and the "rules" of a private language would not be rules at all. A *rule* must be normative: there must be a difference between following it and thinking one is following it. If something is a rule, then it makes sense to speak of "using" it. If it can be "used," it can also be "misused": a rule must allow for the possibility and recognition of mistaken uses. But the "rules" of a private language do not allow for this possibility and recognition. They are arbitrary. Wittgenstein gives two arguments to show this. First, when the solipsist selects a sound (say) as the symbol for a feeling, he has no way of guaranteeing that this association is repeatable, for he has no way to detect his own memory-mistakes. He may think he uses the same word for the same feeling, but he can never be sure. Thus he may think he has established a rule of language, but he has no way of determining whether or not it functions as a rule.²

¹ *Philosophical Investigations*, §§ 65–81.

² *Ibid.*, §§ 258–270.

Second, language is a *public* activity. Necessarily, one's usage of words is subject to correction by other persons. When my right hand gives my left hand money, my financial condition is not changed, and the transaction is a sham. Similarly, when the solipsist gives himself a private definition of a word, he does not create a public instrument of communication. He cannot tell whether he follows his private rule of language or merely thinks he follows it, for there is no one else to tell him, no one to check on his mistakes. As we have seen, an arbitrary "rule" is no rule at all. It justifies nothing. Words have meaning—for words have use—only in a social context.³

The solipsist's picture of "private" sensations depends on an illegitimately extrapolated use of the word "private." "Private" and "public" are contraries, like "fast" and "slow," "light" and "dark." It makes sense to speak of something as "private" only if it might be public: "If as a matter of logic you exclude other people's having something, it loses its sense to say that you have it."⁴ The attempt, then, to find in human experience a pre-linguistic level such as "private" sensations might supply is misdirected. We cannot step outside our language. An "extra-linguistic reality" is a *Ding an sich* whose existence would baffle assertion. The philosopher's problem is not to unveil some external order of fact which could serve as a standard for language, but to understand how language itself functions, to determine what language-games are played. So, briefly, runs Wittgenstein's refutation of solipsism.

I attack this refutation internally. Wittgenstein maintains that linguistic rules are public. This, after all, is what it means to be a "rule." A rule must be a standard of reference accessible in principle to all comers. Accordingly, the use of words referring to other persons involves an implicit recognition of other persons as possible correctors of one's usage of these words, as of any other words. If rules for the use of words are public, then what we may call the "general recognition" of other persons as constituting a public—a public which might correct one's usage—is logically presupposed by the use of words. In particular, it is presupposed by the use of words referring to other persons, that is, by what we may call the "specific recognition" of other persons one grants them by referring to them with names and pronouns. Thus if for purposes of argument we suppose words referring to other per-

sons reducible to descriptions plus the second-person pronoun, then one can meaningfully say "you" only if one is already aware of a community of "you's" or other speakers by which one's usage is regulated. "You learned the concept 'pain' when you learned language,"⁵ that is, when you learned to use the word "pain" as a move in a social game. Having learned to use "you," however, is more than having learned another move in the game: it is the indispensable precondition of playing the game at all, for it supplies the public context with reference to which the game must be played. Wittgenstein is caught in a circle: specific recognition—using "you"—is a move in the game, but playing the game presupposes that the move *has already been made*. Thus Wittgenstein claims to know other persons before he has learned words with which to express his knowledge (of other persons), just as the solipsist claims to know his sensations before he has learned words with which to express his knowledge (of his sensations). The two principles which clash here are:

- (1) "The question is not one of explaining a language-game by means of our experiences, but of noting a language-game."⁶

In other words, it makes no sense to *speak* of an "extra-linguistic reality." We know the world only as we encounter it in language, not beyond language.

- (2) "If I need a justification for using a word, it must also be one for someone else."⁷

In other words, using linguistic rules presupposes that the user recognizes himself to be a member of a linguistic community whose members regulate one another's usage. The conjunction of (1) and (2) is circular, for according to (1), recognition of other speakers is a result of the activity of speaking, while according to (2), the activity of speaking presupposes this same recognition.

II

Discussions of the concept of a "person" tend to center on the substantial problem, "What is a person?" But by concentrating on this particular problem, the philosopher of language assumes without realizing it that a person is a special kind of "thing," so that language referring to persons

³ *Ibid.*, §§ 241, 268 ff.

⁴ *Ibid.*, § 398.

⁵ *Ibid.*, § 384.

⁶ *Ibid.*, § 655.

⁷ *Ibid.*, § 378.

is a special case of language referring to things. He focuses on such questions as "In what ways does it make sense to talk about persons (as opposed to other things)?" and "What predicates can meaningfully be attributed to persons (but not to other things)?"

This is a misdirected emphasis. The philosopher of language should not assume that a person is a special kind of thing. Persons have a function in language-games different from the function of things. We talk *about* things. We talk *to* persons. If one characterizes a thing, the characterization may be corrected by the fellow users of one's language—but never by the thing itself. Persons, on the other hand, can be the literal referents of second-person pronouns and can correct attempts to characterize them. Persons, not things, play the special role in language of *using* language. This is a functional difference. By treating it as a substantial difference—as if a person were a special kind of thing—the philosopher of language is led to treat the concept of a person only within the context of third-person discourse and to neglect the linguistic problems posed by the fact that one speaks not only about persons, but also to them.

These problems cluster around the concept of *recognition*. I shall say that a "person" is whatever may correct one's usage of words, and that "recognition" is the act of conferring on something the privilege of being a person, i.e., a possible corrector of one's usage. Then "general recognition" refers to the implicit acknowledgment of other persons as possible correctors of one's usage which, according to Wittgenstein, is presupposed by the use of language. The users of a language, by using it, collectively legislate what usage shall be "ordinary" or "stock." "General recognition" gives a name to this collective legislating. Language contains words, gestures, and inflections through the use of which a speaker explicitly signifies that he recognizes himself to be addressing an audience. In English, "you" is the most obvious of these. I call the act of using these modes of personal address "specific recognition."

Persons, unlike things, are referents of first-person pronouns. The circle of which I accused Wittgenstein in Section I was generated by arguing that the first-person speaker must know how to use the second-person pronoun "you" in order to place himself in the public context required for using words (including, of course, the second-person pronoun "you"). This is to say that he must be aware of the public nature of language (general

recognition) in order to express his awareness of a public (by specific recognition). But on the other hand, he must have expressed to himself his awareness of a public (i.e., have specifically recognized a public) if he is to be aware of the public nature of language. Thus general and specific recognition are acts each of which presupposes the other. The first-person speaker is awkwardly situated: to be able to speak, he must have provided himself with an audience to speak to, but to provide himself with an audience, he must first speak.

This puzzle has a simple solution. It depends on the assumption that general and specific recognition are the private acts of the first-person speaker. But to assume this is to assume that the speaker is a solipsistic Ego, in other words, that it is meaningful to talk about him apart from the public context in which he expresses himself. This is a false assumption. The solipsistic Ego would require language to articulate its predicament to itself. Since, as Wittgenstein points out, language is public, the predicament is spurious. Recognition, specific or general, is not the private act of the first-person speaker, but a public act in the performance of which, as we shall see, speakers are created.

The solipsistic predicament misrepresents the logic of pronouns. Pronoun declensions are *systematic*. Learning how to use "I" is not different from learning how to use "you," "he," "it," and the rest. The meaning of any pronoun entails the meanings—the uses—of the others. The first-person speaker cannot recognize a requirement for pronominal referents (i.e., for the general recognition of an audience) before referring with pronouns (specific recognition), for the first-person speaker himself is only a pronominal referent. The general recognition of other minds (other speakers) presupposed by using linguistic criteria is prior to the use of words specifically designating other speakers, but prior also to the use of words designating the first-person speaker. In this sense, first persons are not "first."

Thus if one accepts Wittgenstein's assertion that an "extra-linguistic reality" is meaningless, one must apply this assertion to the first-person self as much as to anything else. The self is what one designates when one uses the first-person singular pronoun designatively. This use of language, like any other, requires a public warrant or authorization. This public warrant or authorization is supplied by what I shall call the *mutuality* of recognition. In order meaningfully to refer to oneself in the first-

person singular, one must do so in the presence (actual or implicit) of someone else ("you") who serves as a referee for this usage. "I" is a relation-word which is meaningful only in relation to "you," and "you" is meaningful only in relation to "I." To refer to someone else as "you" (who can correct my usage of "I"), I must confer on him the privilege of referring to himself as "I"—an "I" who may say, "I don't understand you." The designation "I" as applied to me means that I confer on him the privilege of correcting my usage. Therefore, when I confer on him the right to call himself "I," I credit him with conferring on me in turn the right to correct *his* usage. This crediting is meaningless unless in fact he does this. In other words, I cannot autonomously transfer autonomy to him, then back again to myself. Indeed, neither of us is autonomous for neither of us is capable of recognizing himself or the other without the other's connivance. Specific recognition is a public act, a conjoint accomplishment.

Thus the mutuality of recognition makes possible the shifting references of first- and second-person pronouns. A pronoun is only an index—"I" is not the name of anyone—and the recognition granted by the use of pronouns is actually granted to persons as such. But because to be a person is to be able to refer to oneself and others with first- and second-person pronouns, I shall say that to be a person is to play a *role*.

A role is an identity warranted or certified by an audience. The actor on the stage modifies his behavior and appearance so as to seem to be someone else. The recognition of the audience is the criterion of the validity of his performance, that is, of his right to credit himself with his stage identity. Speakers and hearers in ordinary discourse must establish (relatively) permanent identities to make possible the shifting references of pronouns, and they do this by role-playing. "I" means "the one who plays such-and-such a role and who is now speaking," "you" means "the one who plays such-and-such a (different) role and who is now being spoken to." Role-recognition must be mutual, for the speaker speaks in the confidence that the hearer recognizes himself as the person addressed, and the hearer listens in the confidence that the speaker is speaking to him: Indeed, in order to identify himself as "I," the first-person speaker must credit himself with an identity which his hearer can designate as "you," and to do this, he must recog-

nize the hearer as an identity which he, the speaker, can designate as "you." The relativity of pronouns, in other words, is grounded in the relativity of roles. Roles support one another or are defined in relation to one another: the mother must have a child, the ruler subjects, the hero a villain.

Ordinarily we suppose that the recognition of persons ultimately depends on physically locating them in face-to-face confrontations; and in a sense it does. However, while identifying a person's body face-to-face is a necessary condition of recognizing the person, we never treat it as a sufficient condition. *A*: "What is the price of this item, please?" *B*: "I wouldn't know—I'm a customer myself." In a sense, *B* knows he is the person addressed; in the sense, namely, that he knows he is *not* the person addressed and so must tell *A* how to address him. *B*'s role (fellow customer) is the means whereby he makes himself available to *A* as a person, as someone with whom communication is possible, who will respond appropriately to cues and conform within reason to *A*'s expectations. Thus when role-recognition fails to be fully mutual, communication is to that extent impaired, interchanges become irrelevant, and the communicants must reidentify one another until it becomes fully mutual. For example, strangers meet in a smoking-car. *A* hints that he is a celebrity traveling incognito and *B* lets slip that he thinks him an imposter. *A* and *B* then reidentify one another so that *A* takes the role of one whose friendly revelation has been discourteously received, and *B* takes the role of one who is willing to humor an obvious ass. Each recognizes how the other has recast himself. The interchange can now proceed within a new set of limits. But it *can* proceed.

The mutuality of recognition justifies the theatrical metaphor implied by "role." The participants in a linguistic interchange—speaker and hearer, writer and reader—form the audience which collectively passes judgment on role-claims. This audience's specific recognition actualizes the general recognition of one another as possible communicants which the use of language presupposes. In other words, general recognition is a potentiality which exists only in its actualizations, in the multiple and shifting systems of specific recognitions. That it must actualize itself is demonstrated by the fact that when one such system fails to be mutual, the communicants reidentify one another in terms of a system which is mutual.⁸

⁸ I have considered this problem in more detail in "Disagreement as a Dramatic Event," *The Monist*, vol. 49 (1965), pp. 248-261.

The moral is this: language is personificatory, the medium wherein and whereby persons create themselves. Fundamentally, language communicates the self—not as an isolated Ego, but as defined in and by systems of specific recognitions. This personificatory intention is fundamental, for its failure entails the breakdown of communication of any sort. Speech presupposes speakers, who must identify themselves to one another as communicants and define for one another the range of solicitations in terms of which each is accessible as a communicant. Persons, therefore, have a threefold function in language. They are designata: they can

be talked about. They are communicants or role-players: they can speak and be spoken to. They are in their capacity as role-players also role-authorizers: they perform the audiential function of recognizing one another by warranting or accrediting one another's role-claims, collectively creating one another as individual pronoun-designata or persons. In linguistic performances, however, players and audience are the same persons. Therefore the difference in function between player and audience is normally overlooked, and the personificatory intent of language is "veiled" or "self-hypocritical."⁸

Memphis State University

⁸ These are the excellent formulations of Robert Champigny, "Main Intentions in the Use of Language," *The Journal of Philosophy*, vol. 56 (1959), pp. 528-533.

AMERICAN PHILOSOPHICAL QUARTERLY

Edited by
NICHOLAS RESCHER

With the advice and assistance of the Board of Editorial Consultants:

William Alston	Adolf Grünbaum	Wesley C. Salmon
Alan R. Anderson	Carl G. Hempel	George A. Schrader
Kurt Baier	John Hospers	Wilfrid Sellars
Lewis W. Beck	Raymond Klubansky	Alexander Sesonske
Roderick M. Chisholm	Ernan McMullin, S.J.	J. J. C. Smart
L. Jonathan Cohen	Benson Mates	Manley H. Thompson, Jr.
James Collins	John A. Passmore	James F. Thomson
James M. Edie	Günther Patzig	G. H. von Wright
Peter Thomas Geach	Richard H. Popkin	John W. Yolton

VOLUME 3/NUMBER 4

OCTOBER 1966

CONTENTS

I. PAUL D. EISENBERG: <i>Basic Ethical Categories in Kant's Tugendlehre</i>	255	V. G. P. BAKER AND P. M. HACKER: <i>Rules, Definitions, and the Naturalistic Fallacy</i>	299
II. STORRS MCCALL: <i>Temporal Flux</i>	270	VI. RAZIEL ABELSON: <i>Persons, P-Predicates, and Robots</i>	306
III. ALEXANDER SESONKE: <i>Moral Rules and the Generalization Argument</i>	282	VII. JAMES D. WALLACE: <i>Pleasure as an End of Action</i>	312
IV. MORELAND PERKINS: <i>Emotion and the Concept of Behavior</i>	291	VIII. CHARLES CRITTENDEN: <i>Fictional Existence</i>	317

PUBLISHED BY BASIL BLACKWELL WITH THE COOPERATION OF THE UNIVERSITY OF PITTSBURGH

AMERICAN PHILOSOPHICAL QUARTERLY

POLICY

The *American Philosophical Quarterly* welcomes articles by philosophers of any country on any aspect of philosophy, substantive or historical. However, only self-sufficient articles will be published, and not news items, book reviews, critical notices, or "discussion notes."

MANUSCRIPTS

Contributions may be as short as 2,000 words or as long as 25,000. All manuscripts should be typewritten with wide margins, and at least double spacing between the lines. Footnotes should be used sparingly and should be numbered consecutively. They should also be typed with wide margins and double spacing. The original copy, not a carbon, should be submitted; authors should always retain at least one copy of their articles.

COMMUNICATIONS

Articles for publication, and all other editorial communications and enquiries, should be addressed to: The Editor, *American Philosophical Quarterly*, Department of Philosophy, University of Pittsburgh, Pittsburgh, Pennsylvania 15213.

OFFPRINTS

Authors will receive gratis two copies of the issue containing their contribution. Offprints can be purchased at low cost through arrangements made when checking proof.

SUBSCRIPTIONS

The price *per annum* is six dollars for individual subscribers and ten dollars for institutions. Checks and money orders should be made payable to the *American Philosophical Quarterly*. Back issues are sold at the rate of two dollars to individuals, and three dollars to institutions. All correspondence regarding subscriptions and back orders should be addressed directly to the publisher (Basil Blackwell, Broad Street, Oxford, England).



I. FROM THE FORBIDDEN TO THE SUPEREROGATORY: THE BASIC ETHICAL CATEGORIES IN KANT'S *TUGENDLEHRE*

PAUL D. EISENBERG

I

ONE who has read the new translation by Mary J. Gregor of Kant's *Doctrine of Virtue* (*Tugendlehre*),¹ the second part of his *Metaphysics of Morals*, must admit, I think, that this work of Kant's cannot remain little known to the general student of moral philosophy in the English-speaking world; that it demands serious consideration. One need not agree with all or even with much that Kant says there, but one must at least recognize gratefully the abundance of genuine problems which Kant raises there. From this abundance I have selected a few problems to deal with in this paper. These problems concern primarily the conceptual richness of Kant's ethics as revealed in the *Tugendlehre* (but of necessity I shall have to consider briefly some other of Kant's ethical writings in order to clarify the *Tugendlehre* or to set it in proper perspective)—that is, the number of basic ethical categories such as the obligatory and the forbidden that Kant takes account of or allows room for within his actual system of moral philosophy. Important to me also, however, and intimately connected with that concern are the problems raised by Kant's argument establishing obligatory ends, his division of duties into perfect and imperfect ones, and his notion of indeterminate action in pursuit of an obligatory end. How intimately these latter problems are connected with the former will emerge in the course of the paper.

That many traditional ethical theories are insufficiently rich in the number of basic categories that they allow for has been urged in two fairly recent articles by J. O. Urmson² and R. M. Chisholm.³ In his article, Urmson, without citing any texts or offering any close argumentation to support the point, asserts that Kant has no room in his ethics for the concept of supererogation (this part of the assertion I shall show to be correct), but is limited to the inadequate, threefold classification of actions into the morally obligatory, the morally forbidden, and the morally permissible.⁴ And Chisholm accuses Kant of even more grievous error: on the basis of one passage, which he cites in a footnote (the passage, to be considered below, is from Kant's *Religion within the Limits of Reason Alone*), Chisholm maintains that Kant allows for only two categories for actions—the obligatory and the forbidden.

Before attempting to answer these charges one must, of course, determine what are the basic moral categories, some or all of which a given ethical system will deal with. I believe that there are six such basic categories to be distinguished: the obligatory (= right), the forbidden, the permissible and justifiable (= right), the indifferent, the supererogatory, and the offensive.

That these are all the basic categories may, I think, be assumed without further discussion.⁵ What is problematic, rather, is how these categories are to be defined. I shall employ them in

¹ New York, Harper & Row, 1964. Hereafter, especially in citing texts, I shall frequently refer to this translation simply as D.V.

² "Saints and Heroes" in *Essays in Moral Philosophy*, ed. A. I. Melden (Seattle, University of Washington Press, 1958).

³ "Supererogation and Offence," *Ratio*, vol. 5 (1963), pp. 1-14.

⁴ Since I am dealing only with ethics and not also with, say, religion, I shall hereafter omit the explicit qualification "morally" except where confusion would result from such omission.

⁵ In Sect. 6 of the aforementioned essay by Chisholm, an attempt is made to produce a complete listing of the basic ethical categories. Chisholm defines nine such categories; but they seem to be arrived at only by a process of further subdivision of the six categories which I offer here as basic.

accordance with the following definitions⁶: the forbidden is that which one has no right to do (in this and the subsequent definitions the right referred to is a moral rather than a legal right); the obligatory is that which one has no right not to do; the permissible and justifiable, that which one has a right to do or not to do within the area of morality, the area of (greater or less) moral concern—however that area is, or is to be, delimited; the indifferent, what one has a right to do or not to do because what one does is outside the area of morality; the offensive, what one has a right to do or not to do, but which is bad to do or bad not to do; and the supererogatory, what one has a right to do or not to do, but which is good to do or good not to do.

As I intend these terms to be used, no one of them is subordinate to any of the others. But it may be objected that in ordinary usage (or rather, in the ordinary usage of educated men) the indifferent is a subspecies of the permissible. I must agree, in part, with this objection. There does seem to be a notion of the permissible which takes in both morally indifferent actions (such as chewing a piece of gum) and what I may call "morally interesting" actions (such as, to take Sartre's by now famous example, the boy's staying at home, or his going off to war); but I intended my category of the permissible and justifiable to include only morally interesting cases. And, finally, one more note on usage: certain writers use "morally indifferent" as equivalent to "permissible" (in the enlarged sense). Apparently Urmson adopted this use, for on the first page of his essay he refers to "a class of morally indifferent actions, permissible but not enjoined" and cites as an example of actions that are thus permissible "the lead of this or that card at bridge"—an action which is, *per se* and *prima facie*, of no moral interest whatsoever.

We return now to Kant and to Chisholm's charge (contra Urmson) that Kant allows for only two of the six categories which I have distinguished. The passage which Chisholm cites as evidence runs as follows:

It is, however, of great consequence to ethics in general

to avoid admitting, so long as it is possible, of anything morally intermediate, whether in actions (*adiaphora*) or in human characters; for with such ambiguity all maxims are in danger of forfeiting their precision and stability. Those who are partial to this strict mode of thinking are usually called *rigorists* (a name which is intended to carry reproach, but which actually praises); their opposites may be called *latitudinarians*.⁷

Kant appears to be denying in this passage that there are any actions that are permissible in the extended sense—i.e., that are "intermediate" between the obligatory and the forbidden.⁸ This passage, therefore, appears to contradict the *Tugendlehre* both in the latter's threefold classification of actions—

According to categorical imperatives certain actions are *permissible* or *impermissible*, i.e. morally possible or impossible, while some of these actions or their contraries are morally necessary, i.e. obligatory. . . . (D.V., p. 20.)

and in such a passage as the description of the "fantastically virtuous" man

who admits *nothing* morally indifferent (*adiaphora*) and strews all his steps with duties, as with mantraps; it is not indifferent, to him, whether I eat meat or fish, drink beer or wine, supposing that both agree with me. Fantastic virtue is a micrology which, were it admitted into the doctrine of virtue, would turn the sovereignty of virtue into a tyranny. (D.V., p. 71.)

There is, however, no contradiction in fact, just as Mary Gregor avers in her *Laws of Freedom*,⁹ a commentary upon the *Tugendlehre*. She there says briefly¹⁰ that "when Kant denies, in *Religion*, [p.] 23 [of the Prussian Academy edition], that there are morally indifferent actions [i.e., permissible actions in the extended sense] he is referring, not to the objective relation of actions to the law, but rather to the moral attitude of will of the acting subject." Mrs. Gregor's remark is right, but it is not exactly right. It is right because in the passage cited Kant is not concerned with actions *per se* so that there is no contradiction between it and the

⁶ Freely adapted from those offered by John Rawls during a seminar on moral attitudes, at Harvard University, Spring 1964. The crucial terms "(moral) right," "good," and "bad" in these definitions have either to be accepted as primitives or else to be defined in their turn. But stricter definitions than those I offer here do not seem to be necessary for the purposes at hand.

⁷ Kant, *Religion Within the Limits of Reason Alone*, tr. T. M. Greene and H. H. Hudson (New York, Harper & Row, 1960), p. 18.

⁸ That *adiaphora* are intermediate between the obligatory and the forbidden is absolutely clear from the context of the passage cited above. Further, that Kant allows for the categories of the obligatory and the forbidden is manifest in all his ethical writings and no critic has denied this point, which therefore requires no further discussion in this paper.

⁹ Oxford, Basil Blackwell, 1963.

¹⁰ *Ibid.*, p. 108, n. 31.

Tugendlehre's explicit threefold classification of actions, but it is wrong in suggesting (or in not ruling out the suggestion) that Kant is primarily concerned with the attitude of will that a particular moral agent adopts toward a particular action. A fuller quotation from Kant reveals his true meaning:

Between a good and an evil disposition (inner principle of maxims), according to which the morality [as distinguished from the mere rightness or wrongness] of an action must be judged, there is therefore no middle ground. . . . To have a good or an evil disposition as an inborn natural constitution does not here mean that it has not been acquired by the man who harbors it. . . , but rather, that it has not been acquired in time (that he has *always* been good, or evil, *from his youth up*). The disposition, i.e., the ultimate subjective ground of the adoption of maxims, can be one only and applies universally to the whole use of freedom. . . . Further, the man of whom we say, "He is by nature good or evil," is to be understood not as the single individual (for then one man could be considered as good, by nature, another as evil), but as the entire race. . . . (*Religion*, pp. 18–21.)

Kant is thus concerned here with the basic disposition toward evil in Man; and this passage is, therefore, not relevant, as Chisholm presumed it to be, to the question of what categories of moral action Kant allowed in his (purely) ethical writings.

If one discounts as irrelevant to the question before us this passage from the *Religion*, one concludes, on the basis of the text that I have cited from the *Tugendlehre* concerning the threefold division of actions, that Kant means after all to allow for actions that are permissible in the extended sense. But there is still the question whether his purely ethical theory really permits him this category. Thus it can be said and has been said that according to Kant either an action (or more strictly, a maxim of action) violates the categorical imperative, in which case the action is forbidden, or else the action (or maxim) is consistent with the categorical imperative, in which case the action is obligatory, and that there is no other alternative. However, I know of nothing in Kant's writing to support such a view. Rather it seems clear that for Kant the permissible, as the intermediate between the obligatory and the forbidden, is simply that

which satisfies the test of universalizability provided by the categorical imperative (in its first formulation as given in the *Grundlegung*).

How, then, it may be asked, does Kant, by applying this test, determine (in the *Grundlegung*) the duties of developing one's talents and of promoting others' happiness? The answer is that in the *Grundlegung* these duties are established, not by showing that they are consistent with the categorical imperative, but rather by showing that to will their opposites would be to violate the categorical imperative and to fall into self-contradiction. Only in the *Tugendlehre* does Kant use a positive argument—namely, that concerning obligatory ends—to establish these (and other) duties.

Having shown now that Kant not only has, but has in consistency with his basic ethical theory, a notion of the permissible in the extended sense, and having removed, I hope, all obstacles to the view that Kant allows for morally indifferent actions (as indicated by the passage, cited above, concerning fantastic virtue), I wish to determine next whether Kant has a notion of the permissible in the narrower sense (= right₂). Let me consider this question first of all only in connection with what Kant calls perfect duties.¹¹ In connection with such duties Kant *presumably* must allow for right₂ actions for three reasons at least: because of exceptions to rules of duty (so it is commonly said that killing in general is forbidden, but that killing in self-defense is permissible), because of the clash of duty with duty (so it is said by Ross and others that when two *prima facie* duties of equal stringency conflict with one another, it is right₂ to fulfill either one of them), and because of a genuine inability to perform an action required by duty (so it is commonly said that it is permissible for a man who is in ill health not to keep certain promises that he had made beforehand, say, the promise to meet a particular man for luncheon on a particular day; I hope that this example will do, but it must not be supposed that it gives the whole story).

This general presumption may appear unfounded, however. For it may be denied that Kant allows any exceptions to rules of perfect duty; as Mary Gregor observes,¹² Kant is sometimes content to state that laws enjoining perfect duties

¹¹ I shall have a good deal to say later on concerning the distinction that Kant wants to make between perfect and imperfect duties. I prefer not to offer here any definition of "perfect duty"; I presume that the reader has at least a rough notion of what Kant means by that phrase. Incidentally, Kant employs a variety of terms as synonymous with "perfect" and "imperfect" duty. Thus, for example, he sometimes uses the phrases "duties of narrow (or strict) obligation" and "duties of wide obligation" to refer to perfect and imperfect duties, respectively.

¹² *Laws of Freedom*, p. 100.

admit no exceptions. And Kant does say clearly¹³ that there is no such thing as a conflict of duties. Once again, a closer look at what Kant has to say removes the difficulty. I can answer the first part of the present objection no better than by quoting at some length from *Laws of Freedom*:

... on the whole he intends it [the unqualified statement that laws enjoining perfect duties admit no exceptions] to be supplemented by an explanation to the effect that such laws do not admit the moral possibility of arbitrary exceptions, [i.e.,] exceptions made on merely subjective grounds. If a law determines an action precisely, he tells us, "there remains to us no further choice, either for exceptions, if the law is valid in its universality," or for determining what and how much we shall do in observing the law. This suggests that the law may admit certain necessary exceptions or that it may not be applicable under all conditions. Kant does not wish, in fact, to state his position in quite these terms. He criticizes the practice, common in State law, of enacting a prohibition and then listing exceptions to it. . . . What the legislator should do, he suggests, is to include a permissive law [i.e., a law giving a moral title to do something] within the prohibitive law, as a limitation upon the prohibition. His objection to listing exceptions to a prohibition is that this practice indicates that the permissions are given haphazardly or added to the law as particular cases come up. They ought, rather, to be determined according to a principle and so included within the law itself. (Pp. 100-101.)

Thus it appears that, although Kant does not allow exceptions to rules of duty, he does allow exceptions within them; that is, he allows within prohibitions permissive laws, according to which certain actions become right. The second part of the objection can be handled even more easily. While Kant does not allow that duties may conflict with one another, he does allow for what, presumably, is expressed inaccurately in the notion of such conflict—namely, that there can be "two grounds of obligation . . . both present in one agent and in the rule he lays down for himself."¹⁴ Truly, Kant does not go on to make the further point that if these two grounds of obligation are of equal stringency, it is permissible and justifiable for the agent to act upon either one of them. But there is nothing to indicate that Kant would not or could not have

accepted that point if it had been suggested to him; indeed, what else could he have done?

It thus appears that Kant does or must allow for permissible and justifiable actions in connection with perfect duties. What is one to say now with regard to such actions in connection with imperfect duties? There can be no question that the three ways in which right actions arise in connection with perfect duties apply also in connection with imperfect duties—that is, here too there must (because of the very nature of the moral life, if I may be allowed so vague a phrase) be permissive laws within prohibitions, conflicts between or among grounds of obligation, and genuine physical inability to perform one's duty. But if imperfect duties are somehow different from perfect duties, one might well suppose that in that difference there might be found a further way in which right actions would arise in connection with the former sort of duties. In particular, it might be suggested that since Kant, in the special way that I have indicated, allows for necessary exceptions in connection with perfect duties, he would in the *Tugendlehre* allow also for arbitrary exceptions—by which Kant means exceptions made purely on the basis of one's inclinations—in connection with imperfect duties.

Such, in effect, is the argument that Mary Gregor offers in Chapter VII of her *Laws of Freedom*. This argument seems good *a priori*, but Gregor herself admits that her conclusion is not clearly borne out by what Kant actually says in the *Tugendlehre*—not borne out for the reason that in the *Tugendlehre* Kant is writing metaphysics (in his positive sense of that term) so that his "discussion remains on the most general level" whereas "the arbitrary element would show itself in judgment's final application of ethical laws."¹⁵

To try to understand better the role which Kant does assign to arbitrary exceptions in the *Tugendlehre*—as also to decide whether Kant allows in that work for the supererogatory and the offensive—I believe that the distinction between perfect and imperfect duties must be clarified, if possible. Obviously, the questions of whether there may justifiably be arbitrary exceptions to imperfect duties or whether, rather, such exceptions constitute offenses and, finally, whether any action done

¹³ D.V., p. 23.

¹⁴ *Ibid.*

¹⁵ *Laws of Freedom*, p. 106. That reason, however, does not strike me as at all satisfactory; he could well have provided a discussion "on the most general level" on the role of arbitrary exceptions and he could have supplemented that discussion, were it necessary to do so, with a "casuistical note" like those which he in fact appended to various of his general remarks in the *Tugendlehre*.

in performance of an imperfect duty constitutes an act of supererogation all presuppose an understanding of what imperfect duties are and, hence, how they are to be distinguished from perfect duties.

In his early *Lectures on Ethics*, Kant had understood perfect duties to be those to the performance of which one can be compelled by others, and imperfect duties to be those to which only self-constraint is possible. But this simple basis for the distinction was rejected by Kant in the *Grundlegung*—his allowance of perfect duties to oneself is impossible on that basis—and was replaced by a clear-cut, or almost clear-cut, distinction between perfect duties as those duties which allow for no arbitrary exceptions and imperfect duties which do allow for arbitrary exceptions.¹⁶

It is, however, unclear, as I have indicated, whether Kant means to continue in the *Tugendlehre* with this basis for the distinction. In any case, the thought suggests itself that the distinction between arbitrary and no arbitrary exception should not be taken (and should not have been taken by Kant in the *Grundlegung*) as the real basis for the distinction between perfect and imperfect duty, for the former distinction may seem to be at best a consequence of some further and more penetrating distinction that may be used as the true basis of separating perfect from imperfect duties. What else, then, that Kant says in the *Tugendlehre* or elsewhere—and indeed what that he does not mention—may be taken to be that basis?

It might be claimed (on the grounds of what Kant does have to say in the *Grundlegung*) that actions the opposites of which—or, more strictly, the maxims of these opposites—are simply inconsistent when considered as laws of nature are perfect duties. But actions the opposites of which—again, more strictly, the maxims of these opposites—one cannot will in consistency as laws of nature are enjoined by or as imperfect duties. The Kantian doctrine here appealed to is subject to various interpretations, but if it is taken, I think, correctly, to mean that, with regard to perfect

duties, certain actions are ruled out when their maxims are considered as laws of nature because then one would frustrate some end which nature has for one (like the end of self-preservation), then in fact this test of self-frustration is equally applicable in the case of imperfect duties or, rather, their opposites (violations of imperfect duties). On this interpretation Kant's argument in the *Grundlegung* against arbitrary suicide as a violation of a perfect duty (to oneself) is briefly this: nature gives man the desire to avoid pain in order to preserve him, but committing suicide from that desire violates nature's purpose, i.e., goes against the very end which the desire was meant to promote. Similarly, one would like to say that, according to Kant, nature gives man the desire to maintain his self-respect to dispose him to do what the moral law requires.¹⁷ But a man's not showing gratitude in any way to some benefactor of his—Kant takes the showing of gratitude to be an imperfect duty—because he *feels* no gratitude and because he thinks that showing gratitude under those circumstances would be hypocritical and, as such, inconsistent with his respect for himself, violates nature's purpose in giving him the desire for self-respect.

The trouble with this example is only that it is unclear whether Kant regards the feeling of self-respect and the desire to maintain it as natural or not. For on the one hand he says that this feeling and desire are among the "natural dispositions of the mind"¹⁸ and, on the other, that "*consciousness* of these dispositions is not of empirical origin."¹⁹ If Kant's view is that this feeling and desire (as distinct from the consciousness of them) are natural in the same way that the feeling of self-love is natural, then the counter-example may be left as it stands. If, however, Kant's view is that the feeling and desire are not thus natural, then one must say that while Kant did not in fact accept the counter-example, he probably would have, had he been apprised of the present-day doctrine in ego psychology that the feeling of self-respect and the desire to maintain it are purely natural occurrences or dispositions.²⁰

¹⁶ The intended distinction is not wholly clear because Kant is sometimes content to say simply "exceptions" when apparently he means arbitrary exceptions, as his further remarks indicate.

¹⁷ Kant does actually say that self-respect is one of the "natural dispositions of the mind . . . to be affected by concepts of duty—antecedent dispositions on the side of *feeling*" (D.V., p. 59).

¹⁸ *Ibid.*

¹⁹ *Ibid.*; emphasis added.

²⁰ Cf. R. W. White's *Ego and Reality in Psychoanalytic Theory* (New York; International Universities Press, Inc., 1963). Kant would have agreed, I think, with the doctrine in question provided that the self-respect of which he was talking and the self-respect of which, for example, White talks, regardless of the different grounds which those authors give for that respect, are the same feeling experientially.

Leaving this last matter without finally settling it, I turn to the consideration of a way of drawing the distinction between perfect and imperfect duties which is fairly often given. For example, Mill in chap. V of *Utilitarianism* says that "duties of perfect obligation are those duties in virtue of which a correlative *right* resides in some person or persons: duties of imperfect obligation are those moral obligations which do not give birth to any right."²¹ Mill appears to be supported in drawing the distinction in that way by Kant himself. In the *Tugendlehre* Kant says that "only a particular kind of duty, *juridical duty*, [with which, in the *Lectures on Ethics*, he had equated perfect duty], implies corresponding *rights* of other people to exercise compulsion."²² But while he does not explicitly say so, Kant means by "juridical duties" here only outer juridical duties, or perfect duties to others, since he makes it quite clear that no other person has any right to compel someone to perform what Kant sometimes calls "inner juridical duties," i.e., perfect duties to oneself. It must be understood, therefore, that the way of drawing the distinction which is now to be considered is not really Kant's own (at least not Kant's from the time of the *Grundlegung* on).

It must be objected against this (Mill's) proposal as a way of clarifying the distinction which Kant intends to make between perfect and imperfect duties²³ that Kant regards gratitude as an imperfect duty, but one that does involve rights on the part of some other people (namely, one's benefactors). Perhaps those who simply do one a favor have no right to expect gratitude from him; but at least there do seem to be others among one's benefactors who do have that right. For example, a (good) parent clearly has the right to expect that his children show him gratitude in some way in recognition of all that he has done for them. But perhaps gratitude may be regarded as an exception to the general rule for imperfect duties.

Let me take then an example of an imperfect

duty that cannot plausibly be regarded as exceptional, that is indeed often cited as the paradigm of imperfect duties, namely, the duty of beneficence.²⁴ But one may say that even the duty of beneficence involves rights: the rights of humanity at large. Mill himself allows for this move:

... if a moralist attempts, as some have done, to make out that mankind generally, though not any given individual, have a right to all the good we can do them, he at once, by that thesis, includes generosity and beneficence within the category of justice [by which Mill here means to refer to the entire sphere of perfect duties].²⁵

But if (contra Mill) beneficence is still to be regarded as an imperfect duty although it does "give birth to" a right, Mill's distinction must be revised. It must become, not simply a distinction between those duties "in virtue of which a correlative right resides in some person or persons" and those duties with which no such right is correlated (since "some person or persons" can, if only perversely, be construed as including the whole human race in its extension), but rather between perfect duties as involving rights of some specific person or persons (because of certain "special relations" in which they stand to the person who has the perfect duties to them) and imperfect duties as not involving that sort of rights.

Even this revised version will not do, however. For the duty of telling the truth, to give one example, is a perfect duty, and Kant sometimes treats it as a duty which is owed to others. When it is so treated, it surely involves rights on the part of other persons, and not simply some special class of other persons. That is, *prima facie* everyone has the right to have the truth told to him—and this not just in the sense that I must tell the truth to A, with whom I am now speaking, and then to B, with whom I talk later. Rather, if I am (say) an official of the United States government and I am announcing the invention of a new nuclear weapon,

²¹ In *The Philosophy of John Stuart Mill*, ed. M. Cohen (New York, Random House, The Modern Library, 1961), p. 380.

²² D.V., pp. 40-41.

²³ Mill's way of drawing the distinction as one between duties which have rights correlated with them and those which do not is, of course, broader than Kant's distinction between duties with which rights of compulsion are correlated and those with which no such rights are correlated. The latter distinction cannot be taken as the distinction between perfect and imperfect duties that Kant has in mind in the *Tugendlehre*. But the former distinction, while it is not explicitly supported by Kant, does not go against anything that he has to say in the *Tugendlehre* and may, therefore, be taken as a plausible way of elucidating the distinction there between perfect and imperfect duties.

²⁴ Beneficence must here be understood, roughly, as the duty of alleviating or removing the distress of humanity. This understanding corresponds closely to the way in which Kant sometimes defines "beneficence," but it is far removed from, for example, Ross's understanding of it—for which see *The Right and the Good* (Oxford, The Clarendon Press, 1930), p. 24.

²⁵ Mill, *op. cit.*, p. 380.

then, theoretically at least, my words on this occasion may be supposed to reach nearly all of mankind sooner or later; and there is, again *prima facie*, no one in that unspecifiable multitude who has not the right to hear the truth from me. Thus the distinction between perfect and imperfect duties cannot be elucidated in the way proposed by Mill.

Closely connected with this last proposal but still discriminable from it is the proposal that only perfect duties are duties to a specific person or group of persons. (This proposal has nothing directly to say about the *rights* of any persons.) But this proposal is defeated by adapting to it in the obvious way the example, just considered, of the government official and the duty of truth-telling.

Finally, there is the proposal which is most important since it is or is closest to what Kant says most often in the *Tugendlehre* by way of distinguishing perfect from imperfect duties. According to this proposal, perfect duties enjoin or prohibit particular external actions whereas imperfect duties consist only in the adoption of what Kant calls obligatory ends—ends, that is, which it is the duty of man *qua* man (to intend) to promote.

The criticism of this proposal falls into two parts. One must object first of all to the idea that there are some duties (the perfect ones) which require that the moral agent do or refrain from doing certain actions while others (the imperfect duties) require only that he have certain attitudes of will (intentions and motives).²⁶ I believe that there might be duties consisting wholly in the agent's having certain intentions under only three possible conditions, no one of which is in fact realized. That is, there might be such duties (1) if the world were out of joint in such a way that everyone or at least most people could not succeed in carrying out in their actions the intentions which they had or had adopted; or (2) if there were someone who, not because of the constitution of the world at large, but rather because of his peculiar make-up, was a complete moral fumbler so that *he* could never succeed in carrying out the intentions which

he had or had adopted; or (3) if procrastination in carrying out a certain intention were always *morally* possible or permissible. But the unquestionable fact that some people succeed in carrying out their intentions shows that the possibility envisaged under (1) is not realized; and while it is common experience to meet people who are more or less inept, no one, I think, has met a person who could never succeed in doing what he wanted to do. But if all people can carry out some of their intentions some of the time, then it would seem to be evident that it is morally incumbent upon all people to *do* something, sooner or later, to realize their intentions to fulfil what (according to Kant) are and what (still according to Kant) they recognize to be their obligatory ends.

Now suppose that a man, aware of all these facts, recognizes that he must do something sooner or later to carry out such an intention of his, but that he has postponed such action again and again. If he were then to learn somehow that he had only a few more hours to live, would he not (if he remained calm) recognize that he had then to do something to realize such an intention—and would it not then be morally incumbent upon him (to try) to do something? Suppose that he had always intended to show somehow his gratitude to some one but that he had never actually done so although he *had* fulfilled his other duties; would he not then recognize that he had to do, and would he not (*ceteris paribus*) at least try to do, what he had for so long intended to do? Surely the answer to these questions is *Yes*. One may, therefore, assert confidently that if there are certain ends which it is the moral duty of all of us to adopt, then all of us must at some time act toward the realization of these ends. Some external actions in connection with the obligatory ends are morally necessary; otherwise, perhaps, one cannot be properly said to have the *intention* to realize such ends or, alternatively, to have *adopted* such ends. The question then becomes what action one is to do. This brings me to the second part of this final proposal.²⁷

²⁶ I do not insist that this way of drawing the distinction is precisely Kant's own, but I believe that any reader of the *Tugendlehre* will agree that it does at least approximate to what Kant says there. See especially the "Introduction to the Doctrine of Virtue," Sects. I-IX (D.V., pp. 36-56).

²⁷ Perhaps it may be thought that I have dealt unfairly with the proposal so far. It might be said that what that proposal involves is, not the complete separation of external actions and intentions to act (or attitudes of will), but rather mere actions on the one hand as fulfillment of perfect duties and, on the other, some "combination" of attitudes and actions as fulfillment of imperfect duties. According to this understanding of the proposal, one fulfills a perfect duty by doing the right thing, but an imperfect duty by doing the right thing from a (or the) good motive. To this I should object that, depending on the circumstances of the situation or upon our particular frame of mind at a given time, all of us are, or are not, indifferent to attitudes or motives in connection with anything that we regard as a duty at all. In other words, sometimes we take motives into consideration and sometimes we do not; but this difference on our part is not correlated with the difference between imperfect and

In distinguishing perfect from imperfect duties, Kant sometimes appears to mean simply that perfect duties enjoin or prohibit particular actions whereas imperfect duties enjoin indeterminate action in pursuit of an obligatory end (or perhaps it is better to say, enjoin or prohibit a wide range of action). This suggestion cannot be made intelligible, however, until it is understood exactly what a particular action is (or how particular an action must be for Kant to call it "a particular action"). This criticism has been well urged by Chisholm, who writes:

The distinction has . . . been put more broadly: imperfect duties are said to be "indeterminate" in that we have latitude in respect to the manner in which we fulfill them. . . . But if the distinction amounted only to this, then . . . it would require us to say that no duties are perfect. If it is my duty to pay you ten dollars, then I have latitude in that I may pay by cash, check, or money-order; or if it is my duty to pay you in cash, then I may pay by giving you a ten, or fives, or ones; or if it is my duty to give you a ten, then I may give you this one, that one, or the other one; or if it is my duty to give you this one, then I may hand it to you with the face looking up, or down, or right, or left; and so on *ad infinitum*.²⁸

But since the notion of the obligatory end is the leading notion of the *Tugendlehre* and since Kant's doctrine concerning it is especially obscure, it will be worth our while to investigate in much greater detail both it and its connection with Kant's notion of imperfect duty. Let me pause, however, to indicate what, in my opinion, the argument so far concerning perfect and imperfect duties has established and also what I intend to show. Presuming (as I think it is proper to do) that the proposals I have considered are all the *prima facie* likely ways of drawing the distinction between the two sorts of duty,²⁹ or that other *prima facie* likely ways (if there

are any) will prove upon examination to be equally unsatisfactory, I am led to this conclusion: that there is no hard and fast dualism involved here but (at best) a continuum of duties that are of wider or narrower obligation, as Kant's remarks, for example, concerning the relation of duties of respect to duties of love may be taken to indicate.³⁰ But I shall go on to argue that in the *Tugendlehre* Kant presents the reader with more than a continuum of duties; he presents him, rather, with a continuum that ranges from genuine duties to (mere) good deeds and, finally, to saintly and heroic acts of supererogation, although Kant calls this entire range the range of duty.

II

I now return to the discussion of obligatory ends. Kant uses the word "end" in two senses. "In its primary meaning the term 'end' means the intended results or effects of an action. . . . However, Kant includes under 'end' not only the effects we are striving to bring about but also that which serves merely as a limiting condition on our actions."³¹ What is the connection between these two senses? It is, it seems clear, the (real or supposed) *goodness* of the end—a goodness which must be taken account of either positively (in the way that gives rise to the first meaning of "end") or negatively (in the way that gives rise to the second meaning).

Further, Kant recognizes three sorts of ends in themselves or ultimate goods: (a) the good will, (b) humanity (or even rational nature), and (c) the system of the ends of pure practical reason (the *summum bonum*). They are related thus: (b) is the possessor of (a), and (c) the necessary object of (a). Thus (a), the good will, appears to retain a conceptual primacy in Kant's ethics.

perfect duties, respectively. Sometimes, for example, we think it enough if a man gives to charity (in fulfillment of the imperfect duty of beneficence); but sometimes we think, or at least people in general are apt to think, that the man has not *really* done his duty unless he has given out of consideration for the needs of the poor and the thought that it is his duty to help them, and not (say) from the desire to hear himself praised or to see his name in print.

²⁸ Chisholm, *op. cit.*, p. 4.

²⁹ Two proposals that are not based upon what Kant actually has to say may be easily dismissed, as I hope the reader will discover after reflecting on them by himself. These proposals are that only perfect duties have to be done at a particular time and that only imperfect duties are always binding upon a moral agent in the sense that there is never a geographical or locational impossibility to his fulfilling them.

³⁰ Kant says, "although respect is a mere duty of virtue, it is considered *narrow* in comparison with the duty of love, and it is the duty of love that is considered *wide*" (D.V., p. 117). Of course, for Kant both duties of respect and duties of love are imperfect duties, i.e., duties of wide obligation. My point is that the so-called perfect duties are only less wide than the imperfect ones—that there are no duties that are perfect or narrow in the sense of requiring some one absolutely particular action (an action the specification of which has been carried out *ad infinitum*), for particular actions in that sense do not exist. And what other sense is there that would help Kant's argument?

³¹ *Laws of Freedom*, pp. 83–84.

Now Kant maintains that the moral man has the duty to achieve good willing (this is a purely formal characterization of the obligatory end). And that man has a duty to take humanity as his end. Kant states in the second formulation of the categorical imperative in the *Grundlegung*. But the *summum bonum* as such is not an obligatory end for man although it may be said to "contain" obligatory ends, namely, one's own perfection and others' happiness.³² What is left over, i.e., others' perfection and one's own happiness, are not, according to Kant, obligatory ends (although Kant allows that one may have an indirect duty to promote one's own happiness), and he has reasons (whether good or bad, I shall not say) for both exceptions:

Since every man (by virtue of his *natural* impulses) has *his own happiness* as his end, it would be contradictory to consider this an obligatory end. What we will inevitably and spontaneously does not come under the concept of *duty*, which is *necessitation* to an end we adopt reluctantly. . . .

In the same way, it is contradictory to say that I make another person's *perfection* my end and consider myself obligated to promote this. For the *perfection* of another man, as a person, consists precisely in *his own* power to adopt his end in accordance with his own concept of duty; and it is self-contradictory to demand that I do (make it my duty to do) what only the other person himself can do. (D.V., pp. 44-45.)

I have thus constructed one argument—perhaps it might be called the argument *from the summum bonum*—for certain things (one's own perfection and others' happiness) being obligatory ends. But it is not the argument which Kant actually gives in the *Tugendlehre*. His argument there is this:³³

On the basis of earlier writings, we are to assume . . . a categorical imperative for our actions. . . . This first principle of all duty is a principle for actions which, in itself, says nothing about the agent's ends. But human action is purposive, and if the categorical imperative is to be the principle of an action in its entirety . . . then pure practical reason must determine the form of our aims as well as of our actions. . . . Were there no ends of pure practical reason, he argues, then "all ends of practical reason would be valid only as means

to other ends, and a categorical imperative would be impossible—which destroys all moral philosophy" . . . he maintains the purely formal character of the first principle of duty by insisting that the connection between this principle and the ends of pure practical reason is a synthetic connection, and then he connects obligatory ends essentially with our maxims of moral action by offering a "transcendental deduction" in which the nature of human action itself [as purposive] is the "third term" connecting the principle of action and the principle of ends. (*Laus of Freedom*, pp. 85-88.)

Now that we have seen what obligatory ends are for Kant and what arguments he uses or might use to establish certain things as being obligatory ends, it is possible and, I hope, it will be profitable to return to the examination of imperfect duties, which he usually defines as the duties of adopting maxims to promote the obligatory ends. Let me consider in particular what, as I have said, may fairly be taken as the paradigm of imperfect duties, namely, the duty of beneficence, which Kant sometimes defines or interprets as the duty of promoting others' happiness by taking their (permissible) ends as one's own, and sometimes as the duty of helping others in distress, especially financial distress. (That this divergence in definition or interpretation is significant is something that I aim to show.) In this discussion—which, while it is explicitly confined to beneficence, raises issues equally applicable to the case of many, if not all, of the other duties which Kant regards as imperfect—I wish both to clarify his position on several points and to offer criticism of it. That is, I wish to extend Kant's discussion of beneficence in ways which are, surely, consistent with his doctrine but on which he does not enter, or go very far, and then to criticize some of the results that I have produced.

The distress of others which one is to (try to) alleviate or remove is of two sorts, which Kant does not explicitly distinguish. There is, first, distress caused by imminent or present harm or danger; for example, the distress of a man who is aware that he is drowning.³⁴ Now obviously, whether I know the man or not, it is my duty (*prima facie*) to try to

³² Under the heading of self-perfection Kant includes both one's natural and one's moral perfection. In the general duty to promote others' happiness Kant includes both duties of love and duties of respect and also, separately, the duty of friendship. For an elucidation of the content of these various duties and of their interrelations, see *Laus of Freedom*, especially pp. 166-170, 182-184, and 199.

³³ I do not find the following a wholly perspicuous presentation of Kant's argument, but I have not the time here to attempt to clear up various troublesome points in it. On the other hand, this statement of the argument is, I think, not so obscure as only to confuse the reader.

³⁴ I am not at all sure that Kant would really want to classify the duty of helping others in distress of this first sort as imperfect, but there is nothing in his discussion to indicate that he did except it.

rescue him. If I can swim well, then perhaps I ought to dive into the water, get to the man as quickly as possible, and bring him to shore. If I cannot swim well, then perhaps I ought to throw the man a line; or perhaps I ought simply to call others to the scene so that someone among them can help him directly. It is thus true that a variety of actions is open to me. Clearly I must do one or another of these things: from a moral point of view, I cannot arbitrarily decide not to help this man at all, preferring to help someone else at some later time. Nor can I postpone helping this man. Nor do I arbitrarily decide in what way I shall help him; rather, I do whatever I think may best serve him. I therefore have a good reason for what I do and for what I omit to do.

If Kant means by an "arbitrary exception" to an imperfect duty any one of the sorts of arbitrary action just mentioned and if his position in the *Tugendlehre* is interpreted to mean that such arbitrary exceptions are permissible and justifiable, then clearly Kant is wrong. It is more likely, however, that he made no such palpable error and that the suggested interpretation of Kant is, so far at least, a misinterpretation (which, it is true, he does little or nothing to prevent).

I turn now to the second sort of distress, distress caused by long-standing ills or complexes of ills, for example, poverty. It might be said that in this case Kant's notions of the arbitrary exception and of indefinite action in pursuit of an obligatory end are most at home. Purely on the basis of my inclinations I can, it seems, rightly decide whom I shall help and when and how. This view of the matter is exaggerated, however, if it implies, for example, that I can be right in deciding for no good reason, i.e., arbitrarily or on the basis of my inclinations, to help just one of three poor men in my community who have appealed to me (a man economically well off) for financial assistance. The true situation is much more complicated (indeed, so complicated that in the following remarks I shall treat it in an artificially simple way, which I hope will not distort, but rather focus clearly upon, leading features of the true or "real-life" situation). If the three men are equally poor and equally worthy human beings and if I know (somehow) that no one else will or can help them, then it is my duty to help them all equally. Or if I have only enough to give substantial aid to one of them, I must at least give all of them a fair chance of being that one, say, by having them draw straws. So far, there is nothing arbitrary about what I do.

But now suppose that I know somehow or merely have good grounds for believing that others in my community can and will help, say, some pair of them. Then I may rightly choose arbitrarily some one of the men to help and send the others on. Here, indeed, an element of arbitrariness enters; but there is equal opportunity for arbitrary decisions and exceptions in the case of clear-cut perfect duties. Suppose, for example, that I promise personally to give certain information (which, apart from my promise, it would not be even my *prima facie* duty to give) to *some* of those who have asked me for that information (since I know or believe that those whom I have informed will inform the others). Then in all propriety I may arbitrarily choose the exact persons to whom I shall impart the information. Yet Kant treats promise keeping always as a perfect duty; beneficence, as an imperfect duty.

But perhaps there is a situation with regard to my duty of helping the poor which has no parallel in the case of perfect duties. If, say, I have helped a number of poor men in the past, while no one else in my community who could equally well afford to help them did help them, may I not now arbitrarily send on all of a group of persons who have approached me for aid if I know that others will help them once they are approached? I am not at all deciding never to help any of the poor again; I am simply deciding not to help *these* poor. In fact, however, this situation too appears to have its parallel in the sphere of perfect duties. To show that it has, let me schematize the situation thus: I have decided to help neither *A* nor *B* nor *C* now (t_1), but I do intend to help other poor people in the future, perhaps *D* at t_2 , or *E* and *F* together at t_3 , and so on.

Now suppose that I have promised to give interviews to some pressmen and that it would not be my duty to give any interviews except for my promise. Suppose, too, that I do not like giving interviews but that up to now (t_1) I have answered the questions of each member of the press who has approached me. Surely it is quite all right for me to decide arbitrarily that I will not answer the questions of the next pressman who comes to me, whoever he may be; but that after that I will answer the questions of *D* at t_2 , or of *E* and *F* together at t_3 , and so on. I conclude that the notion of the arbitrary exception cannot at all be used to elucidate imperfect duties in distinction from perfect ones.

Finally, the question may be asked: *To what*

extent ought one to pursue, throughout one's lifetime, the duty of beneficence? (Of course, a similar question may be asked concerning any imperfect duty.) But to this question Kant can give no very satisfactory answer. He says that each person must decide for himself to what extent he will perform some imperfect duty; and he suggests, when he is interpreted non-rigoristically (i.e., as holding the rightness of arbitrary exceptions to imperfect duties), that, whatever extent one decides upon or to whatever extent one goes provided only that one does not adopt a maxim of indifference to the duty or adopt a *policy* of minimal and grudging action with regard to it, one will have done one's duty (since, as Kant puts it, the duty consists in the adoption of the obligatory end, not in any particular action or amount of action in pursuit of that end). But Kant does say that the more virtuous a man is, the more he will act in fulfillment of his imperfect duties. While a man thus gains in moral worth if he does much, he does not do that which is forbidden if he does but little (provided, as before, that he has not always intended to do but little).

Kant thus says

To fulfill them [imperfect duties] is *merit* (*meritum* =

+a); but to transgress them is not so much *guilt* (*demeritum* = -a) as rather mere *lack of moral worth* (= 0), unless the agent makes it his principle not to submit to these duties.³⁵

This statement is too sweeping, however, and Kant later qualifies it thus:

To neglect mere duties of love is *lack of virtue* (*peccatum*). But to neglect duty that proceeds from the respect due to every man as such is *vice* (*vitium*).³⁶

This distinction between guilt and simple lack of merit or between the *vitium* and the *peccatum* seems to correspond to Chisholm's distinction (which I have adopted in part)³⁷ between the forbidden and the offensive, respectively.³⁸ The question which I wish to investigate here, then, is what Kant means by "neglect" of an imperfect duty (a duty of love or a duty of respect).

Kant cannot mean total neglect, either in the sense of (1) the agent's adopting a maxim of indifference to these duties, or in the sense of (2) the agent's *not doing anything* in fulfillment of these duties. Kant himself explicitly rules out (1) in the next to the last quotation from him that I have given here; (2) is ruled out by the argument, given in

³⁵ D.V., p. 49.

³⁶ *Ibid.*, p. 134.

³⁷ In his article, Chisholm maintains that offenses may be either minor (trifling) or major (heinous or diabolical). To support his view that there are major offenses, he offers a somewhat complicated example about an informer (for the details of which the reader should consult Sect. 4 of Chisholm's article) and then says that "one might plausibly argue that . . . his [the informer's] act would be permissible but at the same time heinous and inhuman." He does not, however, actually present the argument; and I think that it is not in fact plausible to hold that there are major offenses in the field of morality.

Only minor offenses—which correspond, I believe, to Kant's *peccata*—are, then, permissible forms of behavior from a moral point of view. But the permissibility of such offenses should not be thought to be a distinguishing feature of imperfect duties, for it is easy to see that there are minor offenses committed also in connection with perfect duties. For example, if I have promised (in the ordinary, casual way) to meet someone (for no important purpose) at five o'clock but arrive (not because of unforeseen obstacles, but because of my own "sluggishness") one or two minutes after the hour, I have behaved ill in that I have not strictly kept my promise; and yet what I have done is, under those circumstances, not only permitted, but permissible.

³⁸ The two passages from Kant which I have cited as evidence for my view are really quite confusing. I take Kant to be saying that neglect—which I think must here be interpreted as the making of arbitrary exceptions—of imperfect duties (or, more strictly, duties of love) is ill doing and, as such, opposed to the good, which consists in doing one's duty to the fullest extent that one can, but yet not absolutely opposed to the good as would be *total* neglect of a duty. That the making of such arbitrary exceptions is ill doing seems to be clear from the fact that Kant calls that form of neglect (i.e., non-total neglect) *peccatum*, which is the standard Latin word for sin.

If one focuses, however, not on the word "*peccatum*," but on the word "merit" or "*meritum*," then one may think that Kant means that to fulfill imperfect duties (of love)—which means, perhaps, to carry out in action the maxim that one has adopted to promote the obligatory ends—is supererogatory and, hence, that to neglect so to act by making arbitrary exceptions is not to do anything ill, but only not to do that which is supererogatory.

If the latter interpretation is correct, then I must say that Kant is highly misleading to use the word "*peccatum*" to signify only the not doing of the supererogatory (if that is what sin is, then indeed all men are sinners), and that he is mistaken in attempting an absolute separation between the adoption of a maxim as being one's duty and acting upon that maxim as being supererogatory. For the reasons that I have set forth earlier in the paper, I maintain that if it is one's duty to adopt a certain maxim then it must also be one's duty to act upon that maxim. And if Kant means that to fulfill one's imperfect duties (in this same sense, rather than in the sense of carrying out the maxims) is to perform an act or acts of supererogation (which is what "to fulfill them is *merit*" may mean), then he is committed to the self-contradictory position that something is at once (and without any distinction of aspects or respects) a duty and more or other than duty. My interpretation of the passages, on the other hand, at least appears to be free from contradiction and at best may express clearly what Kant expresses obscurely.

the first part of this paper, that adoption of an obligatory end—an end which it is one's imperfect duty (to intend) to promote—involves (entails, one might almost say) performing some actions in pursuit of that end. Consequently, Kant can mean by "neglect" only the agent's not doing all that he *could* do in pursuit of such an end, which is to say, the agent's making arbitrary exceptions to the self-imposed rule of pursuing and promoting that end.

If this interpretation of Kant is correct, one must conclude that in the *Tugendlehre* he did in fact adopt a position contrary to that which he had held in the *Grundlegung*; i.e., that he no longer believed that arbitrary exceptions in connection with imperfect duties are permissible and justifiable, but believed instead that they constitute permissible ill doing (i.e., are offences). His *peccata* are peccadilloes.

Further, it *might* be thought that if not to do a particular action that one could do in pursuit of an obligatory end, to neglect that end, is *prima facie* an offence, then to do that action is supererogatory—indeed, that anything that one actually *does* in pursuit of an obligatory end beyond the very "adoption of that end" is supererogatory. Of course, to infer that if not doing *A* would be offensive, doing *A* would be supererogatory is a mistake, as Chisholm has pointed out.³⁹ It is, perhaps, a mistake easy to make; but my concern is to determine whether, mistakenly or not, Kant did make the inference which Chisholm has criticized. The passage which, I think, suggests most clearly (or rather, least obscurely) that Kant did make that inference occurs, not in the *Tugendlehre*, but in the brief Second Part of the *Critique of Practical Reason* (written some years before the *Tugendlehre*).

In that part, which concerns the methodology of pure practical reason, Kant is dealing with "the mode in which we can give the laws of pure practical reason *access* to the human mind and *influence* on its maxims,"⁴⁰ and in particular he is concerned to show that more esteem ought to be attached to the strict performance of duty than to the performance of "so-called *noble* (supermeritorious) actions,"⁴¹ i.e., those actions which I have been calling supererogatory. No emphasis should be placed on the "super-" prefix of Kant's "supermeritorious" (in the German, *überverdienstlich*), for soon Kant refers simply to noble and meritorious actions

without apparently intending any distinction between these and supermeritorious actions and certainly without drawing any distinction explicitly. Concerning meritorious or supererogatory actions he writes:

Let us now see, in an example, whether the conception of an action, as a noble and magnanimous one, has more subjective moving power than if the action is conceived merely as duty. . . . The action by which a man endeavours at the greatest peril of life to rescue people from shipwreck, at last losing his life in the attempt, is reckoned on one side [among some people?] as duty, but on the other and for the most part as a meritorious action, but our esteem for it is much weakened by the notion of *duty to himself* which seems in this case to be somewhat infringed. More decisive is the magnanimous sacrifice of life for the safety of one's country; and yet there still remains some scruple whether it is a *perfect* duty to devote one's self to this purpose spontaneously and unbidden. . . .⁴²

The implication of the last words quoted (the words from the semicolon on) seems to be that if the action in question is not the fulfillment of a perfect duty, then it can be only an act of supererogation—so that any action in fulfillment of an imperfect duty becomes an act of supererogation.

But perhaps Kant did not really hold the view that I have attributed to him on the basis of this passage, or perhaps by the time that he came to write the *Tugendlehre* his view had changed. For there what he has to say on the subject of meritorious action is this:

To fulfill the first [division of duties of virtue to others—namely, duties of love] is *meritorious* (in relation to the other person); but to fulfill the second [division of such duties—namely, duties of respect] is to render the other only what is *due* to him. (D.V., p. 115.)

Kant clearly does not say that all action in fulfillment of any imperfect duty is supererogatory or meritorious; nor does he say that action in fulfillment of the imperfect duties of love is simply meritorious. Rather, he says that action in fulfillment of the latter sort of duties is meritorious *in relation to the other person*—which I take to mean, is meritorious from the other person's point of view. In other words, the agent is only doing his duty; but since the beneficiary has, according to Kant, no rights in this sphere, the beneficiary must regard

³⁹ Chisholm, *op. cit.*, p. 6.

⁴⁰ *Critique of Practical Reason*, Second Part, first para.

⁴¹ *Ibid.*, fifth para.

⁴² *Ibid.*, ninth para. My emphasis on "perfect."

the action as meritorious and the agent as doing more or other than duty demands. One cannot, then, say that such action is, without qualification, supererogatory; and perhaps one may think, as I do, that it is better not to speak at all of supererogation in this area since the notion of something that is at once (although in different respects) a duty and an act of supererogation is only apt to confuse. In any case, it is highly misleading to say simply that in the *Tugendlehre* Kant tries to explain the distinction between duty and supererogation as the distinction between perfect and imperfect duty.

While Kant did not, then, intend any part of the *Tugendlehre* to be a discussion of supererogation simply as such, I believe that a fairly large part of that work is taken up with the discussion of actions or modes of conduct which he regards as duties, but which are in fact (major or minor) acts of supererogation. For example, as I have mentioned, he sometimes defines beneficence as the duty of promoting the happiness of others by making their (permissible) ends one's own. This suggests (the suggestion does not seem to be borne out by Kant's examples of beneficent action, however) that it is the duty of each man, not merely to try to remove unjust impediments to others' happiness, but to try to promote their happiness positively—perhaps to try to make them blissful or ecstatic. There is, however, no such duty; in support of this assertion, I offer no argument except the appeal to common opinion. In general, the things which Kant lists as imperfect duties (whether to oneself or to others)—the duties, say, of developing one's mental powers or one's talents, of showing gratitude to the ancients, of cultivating in oneself sympathetic feeling for others, and of not ridiculing even the genuine faults of others—are *not* duties. But to say this is not to say after all that Kant explained the distinction between perfect and imperfect duty as the distinction between duty and supererogation because (1) the distinction between perfect and imperfect duties remains a distinction *within* the class of duties even if some of the examples that he gives of imperfect duties ought not really to be regarded as duties, and because (2) some things also that he lists as perfect duties are not genuine

duties; for example, what he calls man's duty to himself as his own innate judge and the duty not wantonly to destroy the beautiful in inanimate nature.

The error of treating as duties certain things which are properly classed as supererogatory is not, however, an error that Kant is alone in making. Other moral theorists, too, have erred in this regard. For instance, Ross, who certainly did *try* to discern common moral opinion, says that it seems self-evident to him that one has a *prima facie* duty to bring into existence, whether in oneself or in another, the intrinsic goods—namely, virtue, knowledge, and (“with certain limitations”) pleasure—and “to bring as much of them into existence as possible.”⁴³

J. N. Findlay adopts a similar view of the relation between duty and a good or value: we have, he says, a hortatory duty to pursue each value, as also we have a minatory duty to avoid each disvalue.

It is . . . plain that what in a wide sense a man ought to do falls into two quite different segments: a fairly restricted focus consisting of what he is warned off from omitting and therefore ought, in a minatory sense, to do, and a much wider penumbra consisting of the things from whose omission he is not thus warned off, but which he ought, in a hortatory sense, to do. We shall find it convenient to use the word “duty” to cover both segments of what we ought to do. . . .⁴⁴ If we do so, however, not all duty will be “stern,” nor demanding punishment if violated. There will be duties that wear a purely winning aspect, and which will smile, perhaps a trifle wistfully, over a case of omission.⁴⁵

Findlay's view concerning hortatory duty may be summed up in his own words thus: “for every degree of positive value there is a degree of effort which it would be . . . *shameful* not to make, should we know such effort to be probably effective.”⁴⁶ But it is important to note that as an accompaniment of this view Findlay offers a principle of priorities, according to which “the more strongly minatory imperatives . . . by their very nature take precedence over even the most strongly hortatory imperatives. . . .”⁴⁷

Kant's view of the relation between duty and value is, perhaps characteristically, more compli-

⁴³ *The Right and the Good*, p. 24.

⁴⁴ It is to be observed that Findlay is not claiming that his use of “duty” conforms to ordinary usage. I think that his use does not so conform, however.

⁴⁵ J. N. Findlay, *Values and Intentions* (New York, The Macmillan Co., 1961), p. 341.

⁴⁶ *Ibid.*, p. 382.

⁴⁷ *Ibid.*, p. 357.

cated than either Ross's or Findlay's: he requires a "transcendental deduction" to connect the categorical imperative with his doctrine of obligatory ends and with his doctrine that men have a duty to try to promote the *summum bonum*. Whether anything else is wrong with this deduction or not, at least it errs; I believe, in the identification of these obligatory ends. The ends which it is one's duty to adopt are not commonly thought to be the positive promotion of one's own perfection and of the happiness of others, but rather the avoidance of vice in oneself (and, perhaps, in others) and the avoidance of unnecessary suffering on the part of others (and, perhaps, of oneself).⁴⁸

It is well to iterate my grounds for accusing Kant of error here, and for accusing Ross and Findlay of a similar error. The grounds are (1) that those philosophers (among others) have failed to keep to the ordinary understanding of the term "duty" and (2) that as a result their theories leave little or no room for the supererogatory. That both (1) and (2) are errors is a point which, it seems to me, given the present development of linguistic philosophy and its attention to the categories of ordinary thought and speech, no philosopher has any longer to argue for.

It may be well also to indicate (a) how Kant's theory may be changed (but not changed drastically) to accommodate the true view of the supererogatory, and (b) what plea may be entered in defense of Kant's actual position on this topic. On (a): there are two sorts of supererogatory action—what Joel Feinberg⁴⁹ calls "duty-plus" (doing more than is strictly one's duty) and "meritorious non-duty" (doing a good deed that is other than any demanded by duty)—and for both of these Kant in fact leaves no adequate room in his ethical theory as developed in the *Tugendlehre*. But he could easily leave a place for duty-plus by maintaining, with regard to such genuine (imperfect) duties as showing gratitude to one's benefactors and helping others in distress, that there are various minimal extents to which these duties must be fulfilled, and that any action beyond those minima is supererogatory (duty-plus). Just where these minima are to be placed is, no doubt, a difficult ques-

tion to decide; but Kant would not have to decide it in the *Tugendlehre*, for he might hold that it is a question that can be properly decided, not by a metaphysic of morals, but only by the individual judgment of the moral agent in the concrete circumstances of life. And Kant might easily make a place for meritorious non-duty by defining the obligatory ends (in a negativistic way) as the avoidance of vice and of unnecessary suffering (as I have already suggested) so that the positive pursuit of one's own perfection and of others' happiness would become supererogatory.

On (b): while Kant regards much that is supererogatory as a matter of duty, he is at least aware of a difference between this sphere of "duty" and the sphere of genuine duty so that he treats the former, misleadingly, under the heading of imperfect duty and the latter under the separate heading of perfect duty. This separation constitutes a criticism in advance of the utilitarian practice⁵⁰ and an anticipation of Findlay's somewhat better-drawn distinction between hortatory and minatory duties. Further, an incidental remark of Findlay's indicates what I presume to be the basic source of both his and Kant's error. I refer to the remark (already quoted) that "what in a wide sense a man ought to do falls into two . . . segments." This remark points to the source of the difficulty—the word "ought." It is true to say that a man ought to avoid injustice and other moral evils, and it is true to say that he ought to pursue or promote the intrinsic goods so far as he is able to do so. Now what one ought to do to avoid moral evils is one's duty; and it is easy to suppose, as Findlay and Kant do, that whatever else one ought (from a moral point of view) to do is also one's duty.

A further remark of Findlay's suggests another reason (and this time a justifying, rather than a merely explanatory, reason) that a moral philosopher might have for extending beyond its present limits the area of duty. Findlay speaks of the "fairly restricted focus consisting of what [a man] is warned off from omitting" and the "much wider penumbra consisting of the things . . . which he ought, in a hortatory sense, to do." The sphere of

⁴⁸ This wording suggests that it is commonly thought that there are certain ends which it is one's duty to adopt, and it is perhaps wrong to say that *that* is commonly thought. But while the terminology is peculiarly Kantian, it does not, I think, fail to correspond to facts of morality that are commonly acknowledged.

⁴⁹ In "Supererogation and Rules," *Ethics*, vol. 71 (1960-1), pp. 276-288.

⁵⁰ Findlay (*op. cit.*, p. 425) seriously criticizes utilitarianism for having no principle of priorities such as his and for assimilating hortatory duties (the duties of spreading "happiness and other forms of well-being among our fellows") to minatory ones (the duties of not "creating evil or suffering").

minatory imperatives—which, I have argued, is the entire sphere of duty—is indeed narrow, and one may very much want to emphasize the great importance of much that lies beyond that sphere. One way of emphasizing that importance is to treat what is beyond the sphere of minatory duty as included within a larger sphere of (hortatory) duty.

But that way is not a good way since it tends to obliterate what is, if not “one of the deepest gulfs in morals,”⁶¹ at least a boundary line which requires to be observed. Another and a better way—and a way that would be open to the revised version of Kant that I have suggested, under (a)—of emphasizing the importance of much that lies beyond duty would be to say, not that that beyond imposes certain obligatory ends upon rational men, but that it is the subject of justifiable and, perhaps, rationally necessary *ideals*—ideals which it is not one’s duty to try to realize, but which also cannot be treated with indifference by any thinking man or fail to be adopted as *his* ideals.

A brief recapitulation of the main points that I have tried to make may prove helpful to the reader, who has had to consider so many complexities and issues in this paper. I have argued, especially

against Chisholm, that of the six basic categories that a normative ethical theory may recognize and exemplify, five are fairly clearly employed by Kant in the *Tugendlehre*. Concerning the sixth, the category of supererogation, I have shown what little Kant has to say and have suggested the errors that prevented him from saying more (i.e., from giving adequate recognition to that category). To establish those conclusions I have had to investigate in some detail the leading notion of the *Tugendlehre*—namely, the notion of obligatory ends, ends which it is the moral duty of man *qua* man to adopt. Closely connected with that leading notion is Kant’s division of duties into perfect and imperfect ones (which he sometimes defines, roughly, as the duties of adopting and acting upon the obligatory ends). I have considered a number of more or less likely ways of elucidating that division, and I have concluded that that division is not really so sharp as Kant suggests. The sphere of duty comprises a continuum of duties of wider and of narrower obligation; and beyond that relatively small sphere belong many things which one ought, from a moral point of view, to prize and to pursue, but which, contra Kant, do not as such impose duties upon men.

Harvard University

⁶¹ *Ibid.*

II. TEMPORAL FLUX

STORRS McCALL

DOES time flow in the sense that things grow older? What is the difference between past and future? These are the questions to which I shall address myself, and which I hope to go some distance toward answering.

I. THE BLOCK UNIVERSE

I shall begin by presenting a conception of time with which I disagree. Roughly speaking, time on this view is the fourth dimension of a four-dimensional space-time continuum, and spread out in these dimensions are all the things that ever have been, are, or will be, and all the events which have ever happened, are happening, or will happen. The universe, conceived of in this way as a four-dimensional expanse, is correctly and picturesquely describable as the "block" universe. The latter term derives from Bradley, and it is worthwhile quoting the passage in which he uses it:

Or we seem to think that we sit in a boat, and are carried down the stream of time, and that on the banks there is a row of houses with numbers on the doors. And we get out of the boat, and knock at the door of number 19, and, re-entering the boat, then suddenly find ourselves opposite 20, and, having there done the same, we go on to 21. And, all this while, the firm fixed row of the past and future stretches in a block behind us and before us.¹

Or, to change the metaphor, let us conceive of the inhabitants of this world as both animate and inanimate actors in a moving picture that is in the course of being shown. As before, the past and future are spread out in minutest detail on the film, yet we are directly aware only of the present. Shall we say that the present, the flickering screen, is constituted by the *happening* of certain events which otherwise lie locked in the darkened reels? We may

if we wish, but to do so would involve distinguishing between events which are taking place and events which (a) are not taking place, (b) nevertheless in some sense exist, and this distinction surely leads to bad metaphysics. For this reason supporters of the block universe have generally chosen to regard the present as a subjective phenomenon, contingent upon the experiences of conscious beings. Eddington, though not himself friendly to the block universe, puts the matter as follows:

Events do not happen; they are just there, and we come across them. "The formality of taking place" is merely the indication that the observer has on his voyage of exploration passed into the absolute future of the event in question.²

Compare Weyl:

The objective world simply *is*, it does not *happen*. Only to the gaze of my consciousness, crawling upward along the life line of my body, does a section of this world come to life as a fleeting image in space which continuously changes in time.³

It may be objected that these interpretations of the present, based on the Minkowskian picture of individuals as four-dimensional worms in space-time, fall prey to metaphysical infirmities of their own. If events do not happen, but we merely come across them, does our coming across them not constitute a series of happenings? If nothing takes place, except that fleeting images of the world impinge upon our consciousness, why are these successive impingements exceptional?⁴ These objections will not be pursued here, since in this section I am interested only in introducing the block universe. The basic element in this conception is the belief that the categories of past, present, and future have nothing to do with the physical

¹ F. H. Bradley, *The Principles of Logic* (Oxford, 1883), p. 54.

² A. S. Eddington, *Space, Time and Gravitation* (Cambridge, 1920), p. 51.

³ H. Weyl, *Philosophy of Mathematics and Natural Science* (Princeton, 1949), p. 116.

⁴ See G. J. Whitrow, *The Natural Philosophy of Time* (London, 1961), p. 88. Supporters of the block universe might of course reply that "being present to the consciousness of an observer" was just what was *meant* by "happening," so that the former was not itself a happening.

world, though they have everything to do with the observer. This is stated clearly by Russell:

It is of the utmost importance not to confuse time-relations of subject and object with time-relations of object and object; in fact, many of the worst difficulties in the psychology and metaphysics of time have arisen from this confusion. It will be seen that past, present, and future arise from time-relations of subject and object, while earlier and later arise from time-relations of object and object. In a world in which there was no experience there would be no past, present, or future, but there might well be earlier and later.⁵

Here then is the view of time which I shall try to show to be mistaken, or at least to be possessed of a clear and precisely formulable alternative. To give an indication of where my sympathies lie, I conclude this section with a quotation from Broad:

It seems to me that there is an irreducibly characteristic feature of time, which I have called "Absolute Becoming." It must be sharply distinguished from qualitative change, though there is no doubt a connexion between the two. In the experience of a conscious being Absolute Becoming manifests itself as the continual *supersession* of what was the latest phase by a new phase, which will in turn be superseded by another new one. This seems to me to be the rock-bottom peculiarity of time, distinguishing *temporal sequence* from all other instances of one-dimensional order, such as that of points on a line, numbers in order of magnitude, and so on.⁶

II. DETERMINISM

The block universe is *not* necessarily a deterministic one. Nor, conversely, must an indeterministic world be one in which events come into being, rather than one in which they just *are*. Now it is true that some philosophers have denied this. Reichenbach, for example, has attempted to argue that the coming into being of events is implied by the indeterminism of quantum physics. Thus he writes:

The concept of "becoming" acquires significance in physics: the present, which separates the future from the past, is the moment at which that which was undetermined becomes determined, and "becoming" has the same meaning as "becoming determined."⁷

However, as is pointed out by Adolf Grünbaum, there is a confusion here. Strictly speaking, to say that the occurrence of a relatively later event is determined *vis à vis* a set of relatively earlier events, is only to say that there is a functional connection or physical law linking the properties of the later event to those of the earlier events. Hence it is also perfectly proper to speak of an earlier event being determined *vis à vis* a set of later events: it is this in fact which makes retrodiction possible. Now in the block universe we may have partial or even total indeterminacy—there may be no functional connection between earlier and later events. But still, in such a universe, the taking place of an event may consist in nothing more than the entry of its effects into someone's awareness. Of course, in an indeterministic world we cannot, in principle, know what the future will bring. But that does not mean that our eventually coming to know what it brings by waiting and seeing entails that the events in question *happen*. As Grünbaum says:

The belief that in an indeterministic world, the future events come into being or become actual or real with the passage of time would appear to confuse two quite different things: (1) the *epistemological* precipitation of the actual event-properties of future events out of the wider matrix of the possible properties allowed by the quantum-mechanical probabilities, and (2) an *existential* coming into being or becoming actual or real.⁸

Hence the question whether the universe is deterministic is independent of the question whether it is block. It will be argued in the following section that the answer to the second question hinges upon our adopting a certain theory of truth.

III. THE THEORY OF TIMELESS TRUTH

It has been urged by the supporters of the block universe that events do not *happen*, but merely *are*. In this section I shall show that this theory entails what I shall call the theory of timeless truth.

To see what is asserted by the theory of timeless truth, let us consider any true statement in the present tense, such as

(1) Dimitri is eating his soup.

Let us further denote by t_0 the time at which this

⁵ Bertrand Russell, "On the Experience of Time," *The Monist*, vol. 25 (1915), p. 212.

⁶ C. D. Broad, "A Reply to My Critics" in *The Philosophy of C. D. Broad*, ed. P. A. Schilpp (New York, 1959), p. 766.

⁷ H. Reichenbach, "Les fondements logiques de la mécanique des quanta," *Annales de l'Institut Henri Poincaré*, vol. 13 (1953), tr. by A. Grünbaum in *Philosophical Problems of Space and Time* (New York, 1963), p. 320.

⁸ A. Grünbaum, *op. cit.*, p. 324.

statement is made. Then the theory of timeless truth asserts (a) that all assertions like (1) but in the future tense, uttered at times earlier than t_0 , are true, and (b) that all assertions like (1) but in the past tense, uttered at times later than t_0 , are true. It is not necessary for the original present-tense statement to be undated. If instead of (1) we have

(2) Dimitri is eating his soup at t_0

(note that this present-tense statement is false if made at any time other than t_0) then, as before, the theory asserts that

(3) Dimitri will be eating his soup at t_0
is true if uttered earlier than t_0 , and that

(4) Dimitri was eating his soup at t_0
is true if uttered later than t_0 .

Strictly speaking, of course, none of these tensed statements is timelessly true in the sense that its truth is independent of the time at which it is uttered. We might, however, feel inclined to say that the forms of words (2)–(4), uttered at the appropriate times, all expressed one and the same proposition, and that *this* proposition was timelessly true. Following Smart, we might adopt a special device for expressing this proposition, e.g., the italicizing of the verb to indicate that it is to be considered as tenseless.⁹ Then we could write the proposition expressed by (2)–(4) above as follows:

(5) Dimitri *eats* his soup at t_0 .

Here, it might be maintained, we have arrived at a form of statement whose truth is entirely independent of the time at which it is uttered. However, the theory of timeless truth set down here does not require such tenseless forms as (5), although they are useful in that they sum up in a convenient form the common content of (2), (3) and (4), and will be made use of below in Sect. V. All that the theory actually states is that the truth of the assertions (3) and (4) follows, for all appropriate times of utterance, from the truth of (2), or from the truth of (1) uttered at t_0 .

At this point an objection may be raised. Does the timeless theory not imply that certain statements or propositions are true at *certain times*, and does this not involve a logical mistake? Thus, for

example, Waismann maintains that it simply does not make sense to speak of a statement being true or false at a time:

the clause "it is true that—" does not allow of inserting a date. To say of a proposition like "Diamond is pure carbon" that it is true on Christmas Eve would be just as poor a joke as to say that it is true in Paris and not in Timbuctoo.¹⁰

The same sentiments are voiced by Ayer:

So I maintain that if such and such an event is going to happen, then it is true that it is going to happen. It is not true at any special time, whether now or in the future, but just true. To ask *when* it is true is to put an improper question.¹¹

These arguments are persuasive, but they are far from being the last word on the matter. Surely we should allow that if anyone uttered the words "Diamond is pure carbon" on Christmas Eve, he would be making a true statement. At least in *this* sense it seems meaningful to speak of statements being true or false at a time. Hence if the theory of timeless truth seems to license our saying, for example, that the statement "The battle of Waterloo will be fought in 1815" was true in 1814, this is to be taken as implying no more than that if someone made such a statement in 1814, he would be saying something true.

Returning now to the theory of the block universe, it will be seen that this theory implies the theory of timeless truth. For take any event E which does not happen but merely is, say for example, the landing of the first man on the moon. This first man, into whose consciousness (on the block theory) the event E swims, will be entitled to make the true present-tense statement

(6) The (Americans, Russians) are landing first on the moon.

But this immediately entails what is maintained by the timeless theory, namely that the statement

(7) The (Americans, Russians) will land first on the moon

was true before E and at all previous times. For suppose it was not (a Chinese might land first). Then there are no two ways about it, there is

⁹ J. J. C. Smart, *Philosophy and Scientific Realism* (London, 1963), p. 133.

¹⁰ F. Waismann, "How I See Philosophy" in *Contemporary British Philosophy*, third series, ed. H. D. Lewis (London, 1956), p. 457.

¹¹ A. J. Ayer, "Fatalism" in *The Concept of a Person* (London, 1963), p. 237.

no such event as the Russians landing first on the moon, or as the Americans landing first. Similarly, if the following timeless proposition is not true:

- (8) The first men on the moon *return* without mishap,

then there is no such event as their returning without mishap. If there were such an event, then (8) would be true.¹² Hence the theory of the block universe, in which events are but do not happen, entails the theory of timeless truth.

IV. CRITICISM OF THE TIMELESS THEORY

In this section I shall try to show that the theory of timeless truth is false, and hence that the theory of the block universe, which implies it, is false too. Then in the following section I shall offer an alternative theory of truth.

My criticism of the timeless theory takes the form of showing how it is at variance with the ways we ordinarily speak and think about the future and about truth. It falls under three main heads.

(a) *Guessing the future.* Suppose you need some money, go to the local soothsayer, and receive the message that Knobblyknees will win next year's Derby. Suppose that next year Knobblyknees *does* win. Are we to say that the soothsayer's prediction was true from the time he made it right up to the finish of the race? Let us bear in mind that (i) at the finish of the race the present-tense proposition "Knobblyknees wins," is true, and, (ii) to say that "Knobblyknees will win" was true before the race is only to say that if anyone had spoken these words before the race he would have said something true. Then it seems plain that according to the timeless theory, not only the soothsayer's prediction, but also any earlier one, would have been true at the time at which it was made.

In contrast to the timeless view it will be argued here that, on the assumption that the soothsayer was performing nothing more than a sophisticated

act of guessing, his prediction was *not* true at the time he made it, although it would be correct to say that it turned out to be true (i.e., at the finish of the race). This is not, of course, to say that it was false when he made it, but only that it was not true. Ryle makes exactly this point concerning the word "correct," though it is not clear whether he would want to say the same about "true".¹³

To say that someone's guess that Eclipse would win was correct is to say no more than that he guessed that Eclipse would win and Eclipse did win. But can we say in retrospect that his guess, which he made before the race, was already correct before the race? He made the correct guess two days ago, but was his guess correct during those two days? It certainly was not incorrect during those two days, but it does not follow, though it might seem to follow, that it was correct during those two days.

It was stated earlier that though we would be disinclined to say that the soothsayer's prediction was true when he made it, we would perhaps allow that it later turned out to be true, or even, later still, that it turned out to have been true. We could picture ourselves admitting to the soothsayer (ruefully perhaps), "Well, what you said was true after all." But do these words "was true," and "turned out to be true," spoken after the event, carry the implication that what the soothsayer said was true when he said it? Not necessarily. Following Ryle's suggestion in the quoted passage about "was correct," we might interpret the locution

- (1) *X's statement, that it would be the case that p , was true (turned out to be true)*

as meaning no more than the conjunction of (2) and (3):

- (2) *X said that it would be the case that p*

- (3) *It is the case that p .*¹⁴

By studying examples such as these of horse races

¹² We might even convert this proposition and say that if (8) were true, then there would be such an event. In this case the theory of the block universe would not only imply, but be equivalent to, the theory of timeless truth. But such a step is not required for the argument of this paper.

¹³ G. Ryle, *Dilemmas* (Cambridge, 1954), p. 19. At one point Ryle says that guesses are hardly the kind of thing that can be true or false, while at another point, in speaking about prophecies, he says that their being in the end unfulfilled would lead us to *withdraw* the word "true."

¹⁴ This suggestion of Ryle's is followed up by A. N. Prior in *Time and Modality* (Oxford, 1957), p. 94. The slightly different locution:

(4) *X's statement, that it would be the case that p , turned out to have been true,* is properly speaking in order only at times subsequent to and not contemporaneous with the event in question. It carries the meaning of the conjunction of (2) above and,

(5) *It was the case that p .*

we may, I think, weaken within ourselves the temptation to say that every future-tense statement which turns out to be true must have been true at the time at which it was made.¹⁵

(b) *Bringing about the future.* I now want to consider a different type of example, designed to show that statements about the future can occasionally be true. Suppose a doctor, having just given a patient a pill, says

(4) There, *now* he'll get better.

What could these curious words mean? Certainly not that the patient will get better *now*, i.e., immediately. What is implied, I think, is that it is *now true* to say that the patient will get better, whereas it would not have been true to say that a moment or two before. What constitutes the difference? The pill, of course; a new causal factor has been added to the situation which, the doctor's words imply, is sufficient to bring about a cure. This example, as I have analyzed it, contradicts the timeless truth theory; since on that view if the patient gets better, it would have been true to say that he would before the administration of the pill.

Two remarks must be made here. First, as in the case of an incorrect prediction, the statement that the patient will get better is not false before the doctor gives him the pill. If it were false, it would be true then that he would not get better, and this would seem to imply that he would not get better even after the pill. What we seem to be forced to say is that the statement that the patient will get better is neither true nor false before the giving of the pill.

Second, the question must be raised whether it was not true, before the doctor gave the patient the pill, that the doctor would do so. If this question is answered in the affirmative, then it might seem to be true that the patient would get better even before he received the pill, since (i) the pill was efficacious, (ii) it was true that the doctor was going to give it to him. However, from the mere fact that the doctor gave the patient the pill we have no more reason to infer that it was antecedently true that he would do so than we had to believe that the soothsayer's prediction was antecedently true. The doctor, for example, might have decided to do so only at the last moment. Hence we

cannot infer that it was true that the patient would get better before he received the pill. In fact this inference is contraindicated by the original proposition (4).

(c) *The kind of adjective that "true" is.* This final objection to the timeless theory, which derives from Ryle, will not be pressed too strenuously. It takes the form of pointing out that, though "true" and "false" are adjectives, like "sweet" and "white," the properties of trueness and falseness may not characterize statements or propositions in the same way that the properties of sweetness or whiteness characterize lumps of sugar.

As sugar is sweet and white from the moment it comes into existence to the moment when it goes out of existence, so we are tempted to infer, by parity of reasoning, that the trueness or correctness of predictions and guesses must be features or properties which belong all the time to their possessors, whether we can detect their presence in them or not.¹⁶

But this temptation should be resisted. "True" may be an adjective resembling "deceased" or "extinct" more than "sweet." It is characteristic of "deceased" that this adjective applies to persons only after they have ceased to exist, and it may be that "true" also bears an obituary and valedictory aura, as it is clear the adjective "fulfilled" does. But it is difficult to see what are the full implications concerning the notion of truth of these remarks of Ryle's, and no more will be made of them here.

V. A NEW THEORY OF TRUTH

In this section a new theory of truth will be presented as an alternative to the timeless theory. The extent to which this new theory departs from traditional theories (though not, I would maintain, from Tradition in the best sense of the word) is measured not only by its incompatibility with the timeless theory, but also by its incompatibility with Tarski's material criterion of truth. By this criterion, the proposition '*p*' is true if and only if *p*. But if we take for *p* the tenseless proposition

(1) Knobblyknees *wins* the Derby in 1966

¹⁵ A further interesting fact pointing to the same conclusion is that, as a matter of logic, one can neither lie nor tell the truth about the future (see J. L. Austin, *Philosophical Papers* [Oxford, 1961], p. 100). There is of course an exception to this—one can lie about the future when one knows what the future is going to be. For true future-tense propositions (which one can deliberately conceal from others) see (b) below.

¹⁶ Ryle, *op. cit.*, p. 20.

• then the proposition

(2) Statement (1) is true

might be false in 1965 without (1) being false. In this case Tarski's criterion would not be satisfied, and in its most general form it will be rejected in the theory to be proposed here. The following is a summary of the steps to be taken in its presentation.

First, in (a) the terminology of the theory will be restricted, for convenience, so as to deal only with dated tenseless propositions. Then in (b) the time of assertion of a proposition will be distinguished from its time of reference, which in turn will make for an easy method of translation from tenseless to tensed discourse. In (c) conditions will be specified under which tenseless propositions are true at different times of assertion, and finally in (d) the bearing of the new theory on the traditional problem of future contingency will be discussed.

(a) One of the standard criticisms that is made of the block universe is that the type of discourse which it seems to sanction, namely tenseless discourse, in no way does logical justice to our everyday tensed talk. This criticism actually misses the mark, since as we have seen the block universe also sanctions tensed discourse within the confines of the timeless theory of truth. But the feeling remains that, once we confine ourselves to tenseless discourse *à la* Smart, we are saddled with the block universe.

Now I wish to show that this feeling is groundless, and that even if we confine ourselves to tenseless discourse there is logical room for a theory of truth incompatible with the timeless theory, and hence incompatible with the block universe. What kind of universe it is compatible with will be discussed in the last section of the paper. Here we are concerned only with the wedding of precise scientific tenseless discourse to a novel set of truth-conditions.

The theory of truth to be proposed will therefore, as stated, apply only to tenseless dated propositions of the form

(3) The space-time point (x, y, z, t_0) is characterized by property R

¹⁷ We shall, however, briefly consider the problem of more complex tenseless statements in n. 18 below.

¹⁸ Proposition (7), asserted earlier than time t_0 , would seem to correspond to the more complex tenseless proposition

(8) The date of E is some time equal to or greater than t_0 ,

or

(9) $(\exists t) (t \geq t_0 \text{ \& the date of } E \text{ is } t)$.

The truth-conditions of (8) and (9), considered as propositions of the form $p(t_0)$, are the same as those given for propositions of that form in (c) below. Note that no simple tensed statement corresponds to (8) or (9) for assertion times $\geq t_0$.

or

(4) The date of event E is t_0 .¹⁷

But, as we shall see below, it is easy to translate the truth-conditions for tenseless propositions of this form into truth-conditions for tensed propositions. Hence the theory will be applicable to all propositions.

(b) The date t_0 in the proposition (3) or (4), above, we shall call the *time of reference* of that proposition. The time at which it is uttered, written, formulated, cogitated, entertained, or expressed I shall call its *time of assertion*. To every proposition, we shall stipulate for the time being, there may be assigned only one time of reference, though potentially infinitely many times of assertion. It is the distinction between reference and assertion times that makes the translation from tenseless to tensed discourse natural and easy. Thus (3) above, for assertion times before t_0 , will be true just at those times at which the future-tense proposition

(5) The space-time point (x, y, z, t_0) will be characterized by property P

is true, and correspondingly for present- and past-tense propositions. In the event that the time of assertion of (4) is one day earlier than t_0 , (4) is true if and only if

(6) Event E will occur tomorrow

is true at the same time of assertion, and so forth. Tensed propositions of the somewhat different form

(7) Event E will occur,

which are more complicated, will be noted only in passing.¹⁸

(c) We shall now proceed to state necessary and sufficient conditions for the truth of any tenseless dated proposition $p(t_0)$, where t_0 is the time of reference, at different times of assertion. These conditions will be stated in quasi-recursive form; in a way similar to that in which we specify necessary and sufficient conditions for a formula in symbolic logic to be well-formed. The statement of the conditions falls into four parts:

- (i) Proposition $p(t_0)$ is true at assertion time t_0 if $p(t_0)$.

This is the standard Tarski criterion, limited in its applicability.

- (ii) Proposition $p(t_0)$ is true at $t < t_0$ if there exists at t some condition sufficient to make $p(t_0)$ true at t_0 .

The example used in the previous section about the patient will make this clear. If the patient is hovering between life and death, and if $p(t_0)$ states that he is better at t_0 , and if at t he is given a sure-fire pill, then at t there exists a condition sufficient to make $p(t_0)$ true at t_0 , and hence $p(t_0)$ is true at t . In this case it is a causally sufficient condition that is in question, but in other cases it may be a logically sufficient condition. For example, a logically sufficient condition of April 29, 1965 being a Thursday is that April 28, 1965 be a Wednesday. April 28, 1965 is a Wednesday, hence it is true on that day (and also, of course, on earlier days, though doubtless not before the last calendar reform) that April 29, 1965 is a Thursday.

- (iii) If proposition $p(t_0)$ is true at t , then $p(t_0)$ is true at all subsequent times.

This stipulation ensures, for example, that if it were true 2500 years ago that the Greeks discovered the source of the Nile, it would be true today.

- (iv) Under no other conditions (i.e., unless its truth follows from (i)–(iii) above) is $p(t_0)$ true (closure clause).

In parts (i)–(iii) the new theory is in agreement with the theory of timeless truth—only in part (iv) does it differ. According to the latter theory if there are now, 3:30 p.m. April 28, 1965, two pieces of chalk on this table, then when dinosaurs roamed the earth, it was true to say that there would be. The new theory contends that it was not—unless of course (as is unlikely) a Laplacean scheme of universal determinism obtains.

(d) Note that if $p(t_0)$ does not turn out to be true by the above stipulations, this does not entail that $p(t_0)$ is false. In fact a separate and parallel set of

conditions is needed for falsehood—(i) $p(t_0)$ is false at t_0 if not- $p(t_0)$, (ii) $p(t_0)$ is false at $t < t_0$ if there exists some condition sufficient to make $p(t_0)$ false at t_0 , etc. The conditions leave open the possibility (though they do not assert) that there are propositions neither true nor false for assertion times earlier than their times of reference. Such propositions, which we may call future contingents, may become (i.e., be) true or false at later assertion times, but there is no possibility of a proposition changing from true to false or vice versa. Hence the theory of truth proposed here leaves no room for the following interpretation of what may have been in Aristotle's mind when he made his famous remarks about the sea battle:¹⁹

Aristotle may have thought that the truth and falsity of a statement (made at a given moment of time) is determined by its agreement or disagreement with the facts as they are at that particular moment. It is possible that Aristotle's version of the correspondence theory of truth was a theory of *momentary* correspondence. And if so, Aristotle might have thought that the reason the statement "a sea fight will take place tomorrow" is contingent is that its truth value (momentary truth value in the sense mentioned) will still change. At this moment, the admirals are confident and in a fighting mood, and their intelligence underestimates the power of the enemy; in short, the situation is one which naturally leads to a fight. If so, it may be suggested, it will be true to say that there will be a sea fight. But after a couple of hours, the intelligence estimates may have become pessimistic and the admirals timid. The situation presumably will lead to a failure of the sea fight to materialize. If so, then it is perhaps false to say a sea fight will take place.

However, Hintikka's interpretation agrees with the theory of truth propounded here in that the latter, insofar as it concerns future contingents, is describable as one of *momentary correspondence*. A future-tense proposition is neither true nor false if there is not yet anything in the facts for it to correspond or fail to correspond with.²⁰ It is important to note that this theory will not support an epistemic interpretation: future-tense statements are *not* either true or not true depending on what evidence people have for asserting them. It is quite conceivable that many future-tense statements $p(t_0)$ should now be true in the sense that sufficient

¹⁹ J. Hintikka, "The Once and Future Sea Fight," *The Philosophical Review*, vol. 73 (1964), pp. 487–488. The interpretation is put forward in a speculative tone, and Hintikka cites good evidence both for and against its correctness.

²⁰ Compare J. L. Ackrill (tr.), *Aristotle's "Categories" and "De Interpretatione,"* (Oxford, 1963) p. 141 (commentary). Compare also Sect. 9 of J. Lukasiewicz's 1922 paper "On Determinism" (English translation to appear in the forthcoming anthology *Polish Logic*, ed. S. McCall, to be published by the Clarendon Press).

conditions for their truth at t_0 already exist, but that nobody should know this.

Has any philosopher previously maintained a theory of truth similar to the one here proposed? It would seem that at least one has. In Cicero's *De Fato* there are recorded the differing views of Carneades, Cicero's philosophical mentor, and Chrysippus on the matter of determinism. Chrysippus, arguing for determinism, propounds the following argument. First, if there is some movement which has no efficient cause, then it is not the case that every proposition is either true or false. For consider such a movement M . If M has no efficient cause, we cannot say that the proposition that M will occur is either true or false. Second, every proposition is either true or false. Hence every movement has an efficient cause, and is therefore predetermined. Carneades, in reply, agrees with Chrysippus that every movement has a cause, while at the same time maintaining that the voluntary movements of the soul are not caused by anything other than the soul itself. These movements are not predetermined from all eternity, though they have causes. But, Chrysippus rejoins, you have not taken the doctrine that every proposition is true or false seriously enough, for

no proposition about future events can be true, unless it has in the present causes in virtue of which it will be true.²¹

Hence, Chrysippus concludes, all that happens, including the movement of the soul, is predetermined. Plainly Chrysippus is here making use of exactly the theory of truth that is expounded in this section, although his views diverge in the matter of whether every proposition is either true or false. This will be discussed in the next section.

VI. BIVALENCE AND EXCLUDED MIDDLE

It will be apparent that the theory of truth propounded in the previous section, which allows for the possibility of there being propositions neither true nor false for assertion times earlier than their times of reference, is in conflict with the principle of bivalence, which asserts that

Every proposition is either true or false.

However, the theory does not conflict with the law of the excluded middle, which states that two contradictory propositions cannot both be false, or, more succinctly,

Either p or not- p .

The latter is a law of logic which is assertable even in many-valued logics explicitly designed to violate the principle of bivalence, and for this reason the two should be considered as on quite different footings. A paragraph will make this clear.

It was Lukasiewicz who in modern times revived the ancient distinction between bivalence and excluded middle, and his third truth-value, namely "indeterminate," or "possible,"²² joined the familiar "true" and "false" in the construction of a three-valued logic. But although Lukasiewicz makes an explicit distinction between the principles of bivalence and excluded middle in his writings, the waters are muddied slightly by the fact that his system of three-valued logic rejects both these principles. The matrix characterizing his system is the following (where 1 = true, 2 = indeterminate, and 3 = false):

C	1	2	3	N
*1	1	2	3	3
2	1	1	2	2
3	1	1	1	1

and if we define alternation in terms of implication, as Lukasiewicz does,

$$Apq = CCpq, q,$$

we find that the law of the excluded middle $ApNp$ is rejected for $p = 2$. This need not be so, however. If we define Apq , not as $CCpq, q$, but as $CNpq$, then $ApNp$ will be a law of the system.²³ This definition gives, therefore, a system which rejects bivalence and allows excluded middle.

Despite the originality of his approach, Lukasiewicz has done no more; in distinguishing between bivalence and excluded middle, than to revive an ancient Epicurean doctrine. The following quotation, in which Cicero makes the Epicureans blush

²¹ Cicero, *De fato*, xi.

²² Lukasiewicz used the first of these names for the third truth-value (the Polish word means literally "indifferent") in his 1922 paper referred to above, and the second in his "Philosophische Bemerkungen zu mehrwertigen Systemen des Aussagenkalküls," *Comptes rendus des séances de la Société des Sciences et des Lettres de Varsovie, Classe III*, vol. 23 (1930), pp. 51-77 (also to be translated in *Polish Logic*).

²³ Presumably the reason why Lukasiewicz chose the first definition is that the second involves rejecting the law $CAppp$.

for maintaining such an apparent absurdity as to reject one law and accept the other, illustrates this—the Epicurean position, nevertheless, is in harmony with that put forward in this paper.

It is necessary, we say, of two contradictory propositions, and despite the sentiments of Epicurus, that one be true and the other false. Thus from all eternity the proposition "Philoctetes will be wounded" was true, and "He will not be wounded" was false. Unless, of course, we wish to follow the opinion of the Epicureans, who maintain that such propositions are neither true nor false. But they, blushing, proceed even more absurdly to assert that one of a pair of contradictory disjuncts is true, although, considered separately, neither is true.²⁴

It would be difficult to find a more concise statement of such a delicate distinction—all the more remarkable (and all the more likely to be accurate) in coming from such a determined critic as Cicero. A delicate distinction, and yet one that is too often overlooked. To take an example, Steven Cahn believes that if one follows up Richard Taylor's suggestion of denying the law of the excluded middle for future contingents, an absurdity results.²⁵ The argument, couched in terms of the occurrence of a hypothetical sea-battle *SB* at t_2 and the issuing by the naval commander of two orders *O* and *O'* at an earlier time t_1 , proceeds as follows:

- (1) *O* at t_1 is a sufficient condition for *SB* at t_2 .
- (2) Therefore, *SB* at t_2 is a necessary condition for *O* at t_1 .
- (3) *O'* at t_1 is a sufficient condition for no *SB* at t_2 .
- (4) Therefore, no *SB* at t_2 is a necessary condition for *O'* at t_1 .
- (5) But at t_1 it is neither true nor false, according to those who wish to deny the law of the excluded middle (*sic*), that *SB* at t_2 .
- (6) Hence by (2) and (5) a necessary condition for *O* at t_1 is lacking, and by (4) and (5) a necessary condition for *O'* at t_1 is lacking.
- (7) Hence neither order can be issued at t_1 .

The fallacy in Cahn's argument lies in step (6). He no doubt sees that he can achieve his purpose by a straightforward denial of the law of the excluded middle at (5), namely

- (5') Neither *SB* at t_2 nor no *SB* at t_2 .

²⁴ Cicero, *op. cit.*, xvi.

²⁵ S. Cahn, "Fatalistic Arguments," *The Journal of Philosophy*, vol. 61 (1964), pp. 295-305; R. Taylor, "Fatalism," *The Philosophical Review*, vol. 59 (1962), pp. 56-66.

But (5') itself is a contradiction, being equivalent by de Morgan's law to

- (5'') Both no *SB* at t_2 and *SB* at t_2 .

So Cahn instead denies the law of bivalence. However, this denial will not yield (6), for the lack of the truth, at t_1 , of the proposition "*SB* at t_2 ," is not equivalent to the lack of a necessary condition for *O* at t_1 referred to in (2). These difficulties provide another reason for distinguishing the law of bivalence, which can be rejected without absurdity in the sense of having a non-self-contradictory denial, from the law of the excluded middle, which cannot.

The distinction between the two laws will be put to use in Sect. VII. But first, mention having been made in this section of three-valued logic, two very common objections to it must be discussed.

To begin with, it will be said, Lukasiewicz's three-valued system contains all the vices and none of the virtues of two-valued logic. It contains, for example, as laws the paradoxes of material implication *CpCqp* and *CpCNpq*. And it is less simple than two-valued logic. This is true. But these defects of three-valued logic arise not from its three-valuedness but from its truth-functional character; from the fact that the truth-value of complex propositions is determined wholly by the truth-value of their components. It would be perfectly possible to have a logic that was three-valued without being truth-functional.

Secondly, it has been maintained by some logicians that the very idea of a three-valued logic can be explicated only by means of an underlying two-valued logic. Take for example the assertion that in three-valued logic the negation of an indeterminate proposition is likewise indeterminate. Is this assertion true, false . . . or indeterminate? If the last, we would not know what three-valued logic was like. Hence assignments of truth-values in three-valued logic must be given in two-valued logic, and the latter is therefore more basic.

This objection raises some interesting questions. Consider the assertion of the previous section to the effect that:

- (1) Proposition $p(t_0)$ is true at $t < t_0$ if there exists at t some condition sufficient to make $p(t_0)$ true at t_0 .

Is this assertion time-dependent—could there be

assertion times for which it was neither true nor false? No, in this respect it resembles the assertion " $2 + 2 = 4$." But consider the different assertion (2), which also assigns a truth-value to a proposition $p(t_0)$:

- (2) The proposition that the Russians *land* on the moon at t_0 is true at t_0 minus one hour.

Is this assertion true at t_0 minus one year? To say that it was (assuming no sufficient condition then existed) would be to rehabilitate the whole theory of timeless truth, for the truth of (2) at t_0 minus one year would seem to imply the truth of

- (3) The Russians *land* on the moon at t_0

at t_0 minus one year. Hence we must distinguish truth-value-assigning propositions such as (1), which if true at all are truths of logic, from empirical propositions such as (2) and (3) (which may still be truth-value-assigning). It is interesting that the logic of (2) and (3), which is three-valued, must be given a precise form in the language of (1), which is two-valued.

VII. THE ASYMMETRY OF PAST AND FUTURE

Armed with the vital distinction between the law of bivalence and the law of the excluded middle, it will now be possible to give an exact logical characterization of the difference between past and future, and in so doing provide a way of distinguishing the earlier-to-later from the later-to-earlier direction of time. Note that this way of distinguishing the two directions rests upon logic and philosophy rather than upon physics: it is a matter of appealing to the logical status of propositions about past and future rather than to, for example, the increase of entropy or other irreversible processes. In other words, time's arrow points the way it does as a matter of logical necessity, rather than as a contingent matter of fact.

In brief, the difference between past and future is that propositions about the past (i.e., propositions whose times of utterance are later than their times of reference) are subject to the law of bivalence, while propositions about the future are

not.²⁶ Here "subject to the law of bivalence" means that they *must* be either true or false, while propositions about the future are exempt from this necessity. Hence, even if the universe were deterministic, so that every proposition about the future that would one day be true were true now, past and future would still be objectively distinguished. This requires some explanation.

If we reflect upon the theory of truth of Sect. V, which we may call the "temporal" theory, we see that it becomes indistinguishable from the timeless theory, at least insofar as it assigns exactly the same truth-values to all propositions as does that theory, if the universe is deterministic. For what is meant by "deterministic" is that there exist functional connections or physical laws linking the present state of the universe with its state an instant later, of such a type that it would be possible, given Laplace's demon calculator, to predict the future state from the present. It is, perhaps, unlikely that the universe is like this, but if it were, then there would now exist conditions sufficient to make any true future-tense proposition $p(t_0)$ true at t_0 , and hence $p(t_0)$ would be true now, and also by parity of reasoning at any past time. Therefore, if the universe were deterministic, any true tenseless proposition would be timelessly true.

But although this is so, even in a wholly deterministic world the logical difference between past and future would remain. In such a world propositions about the future would all be either true or false, but not as a matter of logical necessity. They would be true or false as a matter of contingent fact, namely as a consequence of the contingent fact that the world was deterministic. (It has never been argued, as far as I know, that the world was deterministic as a matter of logical necessity—those who try to deduce determinism from the law of the excluded middle (a) fail, (b) would succeed, if they succeeded, only in deducing fatalism.) Hence even in a deterministic universe propositions about the future would not be subject to the law of bivalence (though they would in fact be either true or false), whereas, as I shall try to show, propositions about the past both would be and are subject to that law.

²⁶ There is no circularity here arising from the occurrence of the words "later than" in characterizing propositions about the past. It is assumed throughout this paper that the task of ordering the instants of time by the relation "later than" has already been accomplished. That the further task of basing the *directionality* of time (i.e., the asymmetry of past and future) upon its order properties is not a trivial one is indicated by the following quotation from Grünbaum: "Since the instants of anisotropic time are ordered by the relation 'earlier than' no less than by the converse relation 'later than,' the anisotropy of time provides no warrant at all for singling out the 'later than' sense as 'the' direction of time." (A. Grünbaum, "The Anisotropy of Time," *The Monist*, vol. 48 [1964], p. 221.)

Consider any proposition about the past, e.g.,

(1) Socrates *drinks* the hemlock in 399 B.C.

By stipulation (i) of the theory of truth of Sect. V, if Socrates *drinks* the hemlock in 399 B.C., then it is true in 399 B.C. that he does. By stipulation (iii) it is now true. If Socrates does not *drink* the hemlock in 399 B.C. then, by the theory of falsehood of Sect. V, it is false in 399 B.C. that he does. Hence it is now false. But by the law of the excluded middle either Socrates *drinks* the hemlock in 399 B.C. or he does not. Therefore, by the constructive dilemma, it is necessary that either (1) is now true or (1) is now false. That is, (1) is subject to the law of bivalence.

The same line of reasoning will not apply to statements about the future. Consider:

(2) The world *ends* in 2000 A.D.

Even though, if the world *ends* in 2000 A.D., it is true in 2000 A.D. that it does, it does not follow that (2) is now true. Hence the constructive dilemma cannot be applied, and (2) will not be subject to bivalence. In this way a logical difference may be elicited between statements about the past and statements about the future.

VIII. A METAPHYSICAL PICTURE OF TIME

Let us recapitulate the argument to date. The block and the deterministic universes were distinguished, and the former was shown to entail a theory of timeless truth. This theory was criticized, and in Sect. V replaced by temporal theory of truth, which made it possible to differentiate between the laws of bivalence and excluded middle. This distinction in turn provided the basis for a logical asymmetry between past and future. It might be objected that this asymmetry was built into the temporal theory of Sect. V; stipulation (iii) provides that a proposition, once true, is true at all subsequent times, though not at all earlier times. But it might equally well be replied that the *symmetry* of past and future in the block universe (i.e., lack of a "distinguished" direction between the two possibilities "earlier-to-later" and "later-to-earlier") was implicitly built into the theory of timeless truth. In this respect the two theories seem to be on a par.

However, they are not yet on a par in another respect. We find in the Minkowski diagram a clear and perspicuous picture of the block uni-

verse, with all events laid out in their proper spatio-temporal relationships, and observers progressing through space-time as a fence progresses across a field. Can we not construct a different metaphysical picture of the different sort of universe, with its division between past and future, which goes along with the temporal theory of truth?

If we are to construct such a picture, the first thing we must do is to erase, from the Minkowski diagram, all those events which do not correspond to propositions currently true for a given observer. (This immediately makes every picture of the world observer-oriented. The implications of this relativization will be discussed below.) In the world-map of this observer, then, will be found all the events of his (absolute) past, together with those other events for the occurrence of which there currently exist (or existed) sufficient conditions. (In a deterministic world, of course, this would include all events.) What of the remaining events, corresponding to propositions currently neither true nor false? These will all be in the future, or at least outside the absolute past, and for them two possibilities exist.

(a) They can be drawn in sketchily, with outlines fuzzy enough to encompass all possibilities. This seems to be the approach to the future recommended by Peirce in the following passage:

Suppose we open the question of how far the general influences of the theatrical world at present favour the development of female stars rather than of male stars. In order to discuss that, we . . . have to consider the possible or probable stars of the immediate future. We can no longer assign proper names to each. The individual actors to which our discourse now relates become largely merged into general varieties; and their separate identities are partially lost.²⁷

(b) Each set of mutually compatible possible future events (excluding those for the *non-occurrence* of which there exist sufficient conditions) can be drawn on a separate map, thus yielding a picture of the world comprising a sheaf of maps of the future, coupled with a single map of the past. Let us call each of these maps of the future a "possible map." Then the fuzziness of the future posited in alternative (a) may be accounted for by the superimposition of possible maps upon one another, and for this reason I think alternative (b) is preferable.

Here then we have a picture of the universe. How are we to represent the happening of events,

²⁷ C. S. Peirce, *Collected Papers*, 4.172. Quoted in Prior, *op. cit.*, pp. 113-114.

and time's flow? This will involve, in a non-deterministic world, the progressive discarding of those possible maps which do not become actual, and hence part of the map of the past. (There will, of course, always be an infinity of possible maps left.) In a deterministic world progress into the future is represented by the progressive subjection of more and more propositions describing events to the law of bivalence. (There will, of course, always be an infinity of propositions left which are *contingently* either true or false.) In other words, progress into the future consists of the gradual unrolling of the map of the past. This way of conceiving temporal flux has, I admit, a somewhat overly pictorial quality, but I cannot think of a more satisfactory one at present.

It was stated above that the proposed view of time and of truth makes every picture of the universe observer-oriented. This is so, and is an inevitable consequence of that view, but it need not be a bad thing. Let us see how this relativization of world-pictures comes about. Consider two observers O_1 and O_2 located some distance apart, and let there be an event E_1 which is within the absolute past of one but not of the other. Since E_1 is within the absolute past of O_1 , the proposition asserting that E_1 occurred will be true for O_1 . But as E_1 is not within the absolute past of O_2 , O_2 could choose his "now" line, according to the special theory of relativity, in such a way as to place E_1 in the future, relative to him. If there were no conditions sufficient for the occurrence of E_1 in his absolute past, the proposition that E_1 will occur would be neither true nor false for him, and he would be unable to do more than locate E_1 on a possible map.

Stop! It will be said. If E_1 has come within the absolute past of O_1 , must it not *inevitably* one day come within the absolute past of O_2 , assuming that

the two observers are not receding from one another at the speed of light? Can we not assert with confidence that, since the proposition that E_1 occurs is true for O_1 , it must be true for O_2 also, since it certainly will be true for him? This question was discussed in Sect. VI. The answer is *No*, since the proposition

(1) That E_1 occurs is true for O_1

is not at the time either true or false for O_2 . If propositions such as (1) were either true or false for all observers, we would again be confronted with the theory of timeless truth and the block universe. For who is to say that, given any future event whatsoever, there is not some observer (or Observer) for whom the proposition asserting the occurrence of that event is true? As it is, according to the theory put forward here, God could perfectly well cognize every future event, without propositions referring to these events being either true or false in this mundane world.

Given then that world-views, and the truth of propositions, are relative to observers, is this a bad thing? Not necessarily. The concept of absolute or timeless truth suffers at the hands of the theory proposed here a fate precisely analogous to that suffered by the concept of absolute space. Just as in Einstein's theory the results of measuring certain intervals of space and time are relativized to the frame of reference or the observer, so in this theory the truth or falsehood of certain propositions is relativized to their time of assertion. Nor, is it suggested, is the former theory capable of any more precise or scientific formulation than the latter. The flow of time, and the asymmetry of past and future, may yet receive recognition in science and philosophy as characterizations of the physical world as "objective" as the notions of length, date, or velocity.

University of Pittsburgh

and

Makerere University College (Kampala, Uganda)

III. MORAL RULES AND THE GENERALIZATION ARGUMENT

ALEXANDER SESONSKE

OCCASIONALLY, in the context of practical action, someone asks, "What would happen if everyone did that?" Usually, but not always, the questioner seeks to dissuade his hearer from performing some act which he, the speaker, believes wrong. Thus, in some way, the query serves to invoke an argument against the proposed act. The sort of act against which the question is raised is frequently one we have reason to call wrong, yet which is not an offense against any particular individual—such acts, for example, as cheating on taxes, evading such civic duties as jury service, picking flowers in a public garden, etc. We do not normally ask this question about such acts as murder, theft, wanton destruction of property, or needless causing of pain or suffering. About these latter actions we would be more likely to ask, "How would you like it if someone did that to you?" In ordinary moral discourse, then, the initial question seems to evoke a minor moral argument of rather limited application.

In *Generalization in Ethics* (New York, 1961), Marcus Singer assumes from the very first paragraph that anyone who asks, "What would happen if everyone did that?" is using the *generalization argument*, which Singer formulates as: *If everyone did that, the consequences would be disastrous (or undesirable); therefore no one ought to do that.* From the apparently plausible assumption that this particular argument is widely used, Singer moves to more ambitious claims: that the generalization argument serves to establish all moral rules and to determine the relevance of all considerations put forth as reasons for moral judgments. Thus he elevates what appeared to be a minor argument form to a status he identifies as that of a "fundamental moral principle." In this paper I wish to consider whether so exalted a status can plausibly be claimed for this argument.

I

For an argument to be a fundamental moral

principle it must at least, I should suppose, be valid and have a significant and widespread, though not necessarily obvious, role in moral deliberation. As his book proceeds, Singer occasionally cites from the writings of other authors examples which he identifies as "applications" of the generalization argument, thus reinforcing his claim that this argument is widely used in our moral deliberations and disputes.¹ But the reader may have noted, by the time he has reached midpoint in the book, that not one of the examples cited includes an explicit statement of the generalization argument, as Singer formulates it, nor does any of the examples explicitly formulate any argument which is not logically quite different from Singer's generalization argument. He claims, of course, that in each example the generalization argument is *implicitly* involved. But if we make the reasonable assumption that he has used the clearest examples he could find, it is at least interesting that a fundamental moral principle should almost never be explicitly used. This would be a trivial point were it not that the plausibility of Singer's contention that the generalization argument is a fundamental moral principle, depends greatly upon its supposed connection with some of these examples. But if we have reason to think the generalization argument invalid, perhaps we may doubt whether it is even implicitly involved in these examples.

Singer attempts to establish the validity of the generalization argument by deducing it from two "necessary and fundamental principles." This deduction is given in two different forms, which he asserts are both instances of the same deduction "in slightly different language" (p. 66).

In the first formulation, the deduction is:

(1) If the consequences of A's doing x would be undesirable, then A ought not to do x . Singer calls this the *principle of consequences (C)*.

(2) If the consequences of everyone's doing x would be undesirable, then not everyone ought to do x . This is

¹ Such citations occur on pp. 10, 69, 70, 86, 117-118, 135, 167.

called a *generalization from the principle of consequences* (GC).

(3) If not everyone ought to do *x*, then no one ought to do *x*. This is said to be a "form" of the *generalization principle* (GP).

(4) If the consequences of everyone's doing *x* would be undesirable, then no one ought to do *x*. This is, of course, the *generalization argument* (GA), which Singer says follows from (2) and (3) (GC and GP).

In the second formulation of the deduction "ought not" is replaced by "has no right" and the conclusion (GA) has the form: If the consequences of everyone's doing *x* would be undesirable, then no one has the right to do *x*. In this form GA has no initial plausibility whatever and it is not difficult to construct examples which show its invalidity. Singer's claim that the two formulations of GA are equivalent rests upon the propriety of his use of "ought not" and "has no right" as synonyms.² But they are not synonymous in ordinary use; it is not contradictory, or even logically odd, to say, "You ought not do *x*, even though you have a right to do it." It is quite proper to give "you have no right" as a reason for "you ought not," e.g., I ought not enter a restricted military area *because* I have no right to. But "you ought not" is never a reason for "you have no right." The two formulations are not equivalent and I think we may regard this second form of GA as invalid.

The deduction given above of GA in its first, and more plausible, form has been shown by others to be unsound.³ Does this dispose of Singer's claims for GA? Not really, for given the importance Singer accords to GA, it is rather surprising that he should have thought it necessary to provide any proof. For the other two principles which serve as premisses in the deduction, the principle of consequences (C) and the generalization principle (GP), are said to require no proof. Singer says, "If the generalization principle is presupposed in every moral judgment and in all moral reasoning, there is no sense in demanding any further proof of it" (p. 34). Moreover, he also says:

The principle of consequences is a necessary ethical or moral principle. It is necessary not only in the sense that its denial involves self-contradiction. It is necessary also in the sense that like the generalization principle, it is a necessary presupposition or pre-

condition of moral reasoning. There can be sensible and fruitful disagreement about matters within the field delimited by it, but there can be no sensible or fruitful disagreement about the principle itself. We might say that, like the generalization principle, it is both necessary and fundamental. (P. 64.)

Though he uses these principles to prove GA, Singer later says of GA itself:

About the generalization argument, in particular, I shall try to show that it serves to generate and establish moral rules and it is also involved in determining the range of their application, and that, in cases where rules conflict, it serves as the principle for deciding or mediating between them. It is therefore decisive in determining the moral relevance of the considerations that might be advanced as reasons for a moral judgment. (P. 97.)

Moral rules are established by means of the generalization argument. A rule that cannot be derived from an application of the generalization argument cannot be justified. (P. 119.)

What I am urging now is that the generalization argument is also presupposed in every case in which an attempt is made to justify or give reasons for an action. (P. 134.)

Hence it would seem that, having shown GA to be presupposed in all moral reasoning, Singer might well shrug off criticisms of his deduction and claim that the status of GA as a fundamental principle in itself suffices to establish its validity. We must then ask if he has shown that GA is presupposed in all moral reasoning.

II

Singer's claim of fundamental status for GA rests primarily upon the following considerations:

(a) Any genuine moral judgment involves implicit reference to the reasons upon which it is based (p. 55).

(b) Reasons for moral judgments are always instances or applications of moral rules.

(c) All moral rules are established by the generalization argument.

Hence the credibility of the claim that GA is a fundamental moral principle rests primarily upon the arguments which purport to show that GA is always involved in the establishment or justifica-

² The claim that this conforms to general usage occurs in a footnote to p. 65.

³ Cf. David Keyt, "Singer's Generalization Argument," *The Philosophical Review*, vol. 62 (1963), pp. 466-476; G. Ezorsky, review of *Generalization in Ethics*, *The Journal of Philosophy*, vol. 60 (1963), pp. 317-323; J. Howard Sobel, "Generalization Arguments," *Theoria*, vol. 31 (1965), pp. 32-60.

tion of moral rules. It is these arguments which require further examination.

But first let us note the principles which play a major role in all of Singer's discussions of justification, whether of particular actions or of rules.⁴ There are the three acknowledged principles, labeled *C*, *GP*, and *GA* above; i.e., the principle of consequences, the generalization principle, and the generalization argument. But though these are the only principles acknowledged and discussed by Singer, they are not the only ones invoked. For, from p. 83 on, we constantly find a fourth principle in use, i.e., if the consequences of everyone's doing x would not be undesirable, then x is not wrong. This principle occurs in every context in which Singer claims to show that an act can be justified. Since this principle is not given a name or initials, I will christen it *Singer's Justifying Principle (SJP)*. Singer's failure to discuss *SJP* at all suggests that he believes that it either follows from *GA* or is another alternate form of *GA*, and that by showing that *SJP* is involved in justification, he is thereby showing that *GA* is involved. The relevant form of *GA* here would be: If the consequences of everyone's doing x would be undesirable, then x is wrong. But we all know that " $\sim p \supset \sim q$ " neither follows from nor is equivalent to " $p \supset q$ "; yet this seems to be the relation of *SJP* to *GA*.

SJP would follow from *GA* if *GA* were stated: x is wrong, if and only if the consequences of everyone's doing x would be undesirable. But it would be plausible to assert this equivalence only if either of two conditions held:

(1) Singer's system contained only one fundamental principle, *GA*. If all other principles and rules were derived from *GA*, then any act to which *GA* did not apply would not be wrong. But *GA* is not Singer's only principle; in fact he purports to deduce *GA* from *C* and *GP*; or

(2) Though there are a number of principles, it is impossible for them to conflict. Then, if *GA* were relevant in every moral situation, as Singer claims, any action which was not shown to be wrong by *GA* could not be shown to be wrong by any other principle either, and we could assert *SJP*: if the consequences of everyone's doing x would not be undesirable, then x is not wrong.

Singer does claim that condition (2) holds, and though he does not say so, we may presume that it is upon this basis that he uses *SJP* so freely. On p. 105 he lists a number of moral principles and then considers the possibility of conflict among them. I shall quote all of his argument:

A little reflection suffices to show that it is impossible for any of these principles to conflict, though they are all closely related, and this is a further important difference between moral rules and principles.

Yet it might be supposed that there is a possibility of conflict between the generalization argument and the principle of consequences. For the consequences of an action in a particular case might be undesirable, while the consequences of the general performance of that sort of action might not be undesirable, and this would seem to give us incompatible results. So it is advisable to examine this possibility.

Suppose, then, that if A were to do x the consequences would be undesirable; it follows, on the principle of consequences, that A ought not to do x . Suppose, also, that if no one were to do x , the consequences would be undesirable; it would seem to follow, on the basis of the generalization argument, that everyone, including A , ought to do x . But there are two possibilities here: (1) if everyone were to do x , the consequences would also be undesirable; or (2) if everyone were to do x , the consequences would not be undesirable. In the first case the generalization argument is invertible⁵ and nothing follows from it. So in this case the conclusion from the principle of consequences prevails, and A ought not to do x . In the second case, there must be something distinctive about A , or the circumstances in which he is placed, to explain the difference between the consequences of his performing an action and the consequences of everyone's performing it. If everyone similar to A did x in a similar situation, the consequences would be undesirable, and so in this case also it follows that A ought not to do x .

I conclude from this that there is no possibility of a conflict between these two principles, and hence that there is no possibility of a conflict between any of them.

But does the argument support this conclusion? Consider the following example, which is not merely a logical possibility. Let A be an adult Negro living in Mississippi in, say, 1955; and let x be the act of seeking to register as a voter. If A , alone, were to do x , the probable consequences would be undesirable. The consequences for A

⁴ Cf. Chapter V, also pp. 83-85. For the purposes of this paper I shall not question Singer's claims about either *C* or *GP*. My concern is rather to show that, granting *C*, the claims made for *GA* are false. I have also adopted Singer's terminology in talking about rules and principles, though I do not think it is the clearest terminology to use.

⁵ Singer calls *GA* invertible when the consequences of everyone's doing x would be undesirable, and the consequences of no one's doing x would also be undesirable. If *GA* is invertible it is invalid, and, Singer says, nothing follows from it.

might be that he be threatened, beaten, lose his job and his credit, be arrested on false charges, have his house burned or bombed, etc. The hostility of whites toward Negroes in *A*'s community would be increased, while Negroes would gain nothing at all from *A*'s action. It follows, from the principle of consequences, that *A* ought not to do *x*. But if no one similar to *A* were to do *x*, the consequences would also be undesirable. If no adult Mississippi Negro sought to register, then the Negroes of Mississippi would continue to be deprived of civil rights and to be depressed and abused. However, as Negroes have begun to realize in more recent years if *everyone* did *x*, the consequences would not be undesirable. If every adult Mississippi Negro sought to register, then the system of suppression would crumble and the period of Negro servitude would be near its end. Hence it would follow from the generalization argument that every adult Negro, including *A*, ought to do *x*. Here *C* and *GA* clearly conflict. What Singer fails to consider is that there are cases where the decisive difference which determines whether the consequences would be desirable or not is just the difference between a single person acting and everyone doing the same act. There is nothing about *A*'s act which accounts for the difference in consequences, except that everyone does not do *x*. Thus *C* and *GA* can conflict, and *SJP* does not follow from *GA*, or from any of Singer's principles. It is in fact an independent principle for which he has provided no justification. Hence Singer cannot properly use occurrences of *SJP* in support of his claim for the fundamental character of *GA*.

III

We may now turn to those sections in which Singer argues that moral rules, and therefore moral reasons, presuppose *GA*. His most extravagant claim occurs on p. 119: A rule that cannot be derived from an application of the generalization argument cannot be justified.

Singer divides moral rules into three types. These are: (1) "neutral norms," e.g., the rules of the road; (2) "local rules," e.g., the rule that everyone ought to pay his taxes; and (3) "fundamental moral rules," e.g., rules against lying, stealing, or killing (p. 112). Our question is, has he shown that rules of all three kinds are "established by means of the generalization argument?"

First, consider neutral norms. A neutral norm

is called such because "it would make no moral difference if its opposite were adopted" (p. 113). The example Singer uses here is the rule that everyone is to drive on the right hand side of the road. He writes:

The characteristic of a neutral norm is that the same results would have been attained by adopting precisely the opposite, while it is necessary to adopt *some* rule. . . . That it is necessary to have some rule is established by the application of the generalization argument. What would happen if there were no rules for directing and ordering traffic, if everyone drove on the same side of the road, or on the side of the road on which he happened to feel like driving? There is no need to specify the details. It is surely clear that this would be extremely inconvenient, and this is sufficient to show that not everyone ought to drive on the same side of the road, and that no one has the right to drive on the side of the road on which he happens, at the moment, to feel like driving. It follows that everyone ought to drive on the right, or on the left, whichever is in accordance with the rule of the community whose roads he is using; unless he has, as he might in special circumstances, good reason for the contrary. It also follows that there must be some rule to prevent catastrophe and serve the needs roads were built to serve. (P. 114.)

But here, as elsewhere, Singer's failure to specify the details conceals the flaws in his argument. For if we do specify the details it is clear that the generalization argument is not applicable here at all. Consider each of the questions assumed to be examples of *GA*. (1) What would happen if there were no rules for directing and ordering traffic? This does not even have the form of what we may call the *generalization question*, what would happen if everyone did *x*? and certainly does not lead to any conclusion of the form found in *GA*, no one ought to do *x*. If we carry out the argument suggested by (1), it would appear to be that the consequences of having no rules for ordering traffic would be undesirable; therefore there ought to be rules for ordering traffic. But this is not *GA*; rather it is a straightforward application of *C*, the principle of consequences. This is more apparent if we formulate the argument as: If the community does not establish rules of the road, the consequences will be undesirable; therefore the community ought to establish rules of the road. The second question Singer asks is (2): What would happen if everyone drove on the same side of the road? This does have the form of the generalization question, but if we attempt to carry out *GA* from this ques-

tion we rapidly degenerate into nonsense. The premiss, the consequences of everyone driving on the same side of the road would be undesirable, seems sensible and true; but the conclusion, therefore no one ought to drive on the same side of the road, is either completely incomprehensible or obviously false. This can be seen if we ask if this instance of *GA* is invertible. If we ask, "What would happen if no one drove on the same side of the road?" we must either admit that the question is nonsense, or if we insist that it is intelligible, and means just what it says, then the consequences would be undesirable; for they would seem to be that we all give up driving. Hence question (2), construed as an application of *GA*, is either nonsense or invertible. In either case it does not serve to establish any moral rule. The final question in this cluster, (3): What would happen if everyone drove on the side of the road on which he happens to feel like driving? This again invites application of *GA*. But, again, if it is, it does not establish any moral rule; for the argument is clearly invertible. If *no one* drove on the side of the road on which he felt like driving, the consequences would be identical with those which would ensue if everyone did, and equally undesirable. Hence nothing follows from the argument; certainly not Singer's conclusion that no one has the right to drive on the side of the road on which he feels like driving. The actual situation is that if there are no rules we all have the right to drive where we like, as we all have the right to eat what food we like. The need for rules is established by *C*, and once rules are established and we are taught to drive in accord with them, then most of us most of the time will in fact feel like driving on that side of the road specified by the rules. And it would then be ludicrous to say that we have no right to do so.

Thus, of the questions asked in purportedly establishing this neutral norm, two are either nonsense or invertible when understood as applications of *GA*; while the third, which does establish the need for rules, is an application of *C*, not of *GA*. And once the necessity of having some rules is established by *C*, the determination of particular rules does not require an application of *GA* either. If we are considering alternative sets of rules, *F*, *G*, and *H*, then surely the reasonable procedure is to establish those rules whose adoption will have the most desirable consequences and will least conflict with existing rules and practices, i.e., that set of

rules which will best serve the needs of the community.

We may then conclude that neutral norms are not derived from or established by *GA*, or, minimally, that Singer has not shown that they are.

IV

The second class of rules, which Singer calls "local rules," are peculiar to different groups or communities. The primary example cited to show that such rules are "derived" from *GA*, is the following:

"Two systems of water law are in force within the United States—the riparian and the appropriation systems." The system first named prevails in thirty-one of the forty-eight states. Its fundamental principle is "that each riparian proprietor has an equal right to make a reasonable use of the waters of the stream, subject to the equal right of the other riparian proprietors to make a reasonable use." Some of the arid states of the west found this system unsuited to their needs. Divisions of the water "into small quantities among the various users and on the general principle of equality of right" would be a division "so minute as not to be of advantage to anybody." "It is better in such a region that some have enough and others go without, than that the division should be so minute as to be of no economic value." The appropriation system is built upon the recognition of this truth. Its fundamental principle is "that the water user who first puts to beneficial use the water of a stream, acquires thereby the first right to the water, to the extent reasonably necessary to his use, and that he who is the second to put the waters of the stream to beneficial use, acquires the second right, a right similar to the first, but subordinate thereto, and he who is the third to put it to use acquires the third right, a right subordinate to the other two, and so on throughout the entire series of users."⁶

Is *GA* involved in this example, as Singer claims? I think not. If *GA* were applied, its form would be: If every riparian proprietor used as much water as he needs, the consequences would be undesirable; therefore no riparian proprietor ought to use as much water as he needs. But no such argument is mentioned, or evoked. And rightly so; for the argument in this form is clearly invertible—the consequences of no one using as much water as he needs would also be undesirable, as Singer notes. But if *GA* is invertible, it is invalid. And though Singer says that if *GA* is invertible "nothing follows from it," here we find him claiming that a

⁶ Singer, *op. cit.*, pp. 117–118. The example is taken from B. N. Cardozo, *The Growth of the Law* (New Haven, 1924), pp. 118–119.

local rule is derived from an invertible (and invalid) instance of *GA*.

If we reconstruct the reasoning which supports the particular local rule, it is clear that *GA* is not involved at all. What is involved is not *GA*, but the *premiss* of *GA*, which we might call the *generalization premiss*: if everyone did *x*, the consequences would be undesirable. Singer wrongly assumes that every time this premiss occurs, the whole generalization argument, with its conclusion, therefore no one ought to do *x*, is evoked. But the conclusion which does follow from the generalization premiss is not this, but rather "therefore it ought not to be the case that everyone does *x*,"⁷ here, it ought not to be the case that every riparian proprietor uses as much water as he needs. Here the generalization premiss is not used as the premiss of *GA*; rather the whole argument is an instance of the principle of consequences. When the principle of consequences, in this form, applies to a situation, and no fundamental moral rules also apply, a new consideration becomes relevant, i.e., is it likely that everyone, or most everyone, will do *x* (or will want to do *x*)? Since it is likely that all riparian proprietors will want to use large quantities of water, the conclusion reached is that there ought to be rules controlling the use of water in the area. The particular rules to be adopted are determined by considerations of fairness and the desirability of probable consequences.

Here, as in his discussion of neutral norms, Singer has failed to show that *GA* is involved in any way in the establishment of the rules. Again he has mistaken a use of the generalization *premiss* for an application of the generalization *argument*, but in these examples the premiss functions in an instance of the principle of consequences, not of *GA*. We may note another facet of this mistake. Singer argues, unconvincingly, that the reply, "But not everyone will do *x*" is irrelevant to the generalization argument. Perhaps so, but as we have seen, it is not irrelevant to a generalized instance of the principle of consequences. If our conclusion is "It ought not to be the case that everyone does *x*," the

response, "It will not be the case that everyone does" (not everyone will do *x*), is both relevant and sufficient, if true, *as long as no other moral considerations are involved*. No further action need be taken; no one need be denied the right to do *x*. For example, it seems obvious that if everyone were to refuse to drink milk the consequences would be undesirable, both economically and in terms of individual health. But it is equally obvious that not everyone will refuse; hence we need no laws requiring milk consumption, nor need we limit anyone's right to refuse to drink milk *without giving any reason*. It is only where it is likely that all (or most) people will do *x* that the truth of the generalization premiss justifies the imposition of rules limiting individual rights.⁸ Thus we might say that when there are no reasons against a type of action except the generalization premiss, the response, "Not everyone will do *x*," if true, *neutralizes* the generalization premiss, i.e., no significant moral conclusion follows from it.

Thus Singer has not shown that *GA* is involved in establishing either neutral norms or local rules. But perhaps a different question is really at stake, though not clearly stated. Perhaps Singer's claim is not that rules are derived from *GA*, but rather that *GA provides our reason for obeying the rules*. That is, given a neutral norm or local rule, *N*, if we ask, "Why should I act in accord with *N*?" the decisive answer is "If everyone disobeyed, the consequences would be undesirable; therefore no one ought to disobey." But now we are on quite different ground, the principle "one ought to obey the laws of the community" is neither a neutral norm nor a local law, but, in Singer's terms, a fundamental moral rule. Hence I shall not consider this question until after asking whether Singer shows that fundamental moral rules are derived from *GA*.

V

In his discussion of fundamental moral rules Singer does not cite any purported example of the derivation of such a rule from *GA* by another

⁷ Cf. the n. to p. 65, where Singer suggests this as an alternative formulation of "not everyone ought to do *x*." Unfortunately, Singer shows no apprehension of the difference between these formulations. But there is an important difference. In "it ought not to be the case that everyone does *x*" the *ought* is clearly evaluative, an ought, (see my *Value and Obligation*, New York, 1964) and nothing follows from it about rights or obligations. But in "not everyone ought to do *x*" the *ought* is ambiguous, and might be mistaken for an ought—the ought of obligation.

⁸ But note that when there are rules prohibiting *x*, one cannot justify disobeying the rule by saying, "Not everyone will do *x*." The *propriety* of the rule is quite independent of the truth of the generalization premiss, though in some cases the *wisdom* of the rule may not be. Singer completely fails to distinguish between the use of this premiss in situations where there are rules and in situations where no rules are in force. He must ignore the distinction, of course, because of his contention that all rules are derived from *GA* and his failure to distinguish uses of *GA* from uses of the premiss in other argument forms.

author, but sets out to show himself how a particular fundamental moral rule is established. His argument is the following:

Let us turn to the question of justifying fundamental moral rules. The procedure, as already indicated, is the same in every case. Moral rules are established by means of the generalization argument. A rule that cannot be derived from an application of the generalization argument cannot be justified.

Since the procedure in every case is the same, it does not matter which rule we select to exemplify it. Let us take the rule that lying is wrong. What is the proof of this? Since to justify a moral rule is equivalent to explaining why a certain kind of action is generally right or wrong, to justify the rule against lying is equivalent to explaining why lying is wrong. Thus it will be sufficient to answer the question "Why is it wrong to lie?" . . .

The reason lying is wrong should be obvious from what has already been said. Lying is wrong because of what would happen if everyone lied. It would be nothing short of disastrous if everyone were to lie whenever he wished to, if lying became the rule and truth-telling the exception, which is, however it may seem, actually not the prevailing practice. It follows that lying is generally wrong, or that no one has the right to lie without a reason, and that the mere wish or desire to lie is never a sufficient justification. (Pp. 119, 121.)

The weakness of this argument is so apparent that it is surprising Singer should allow so much to rest on it. The claim that it is obvious that the consequences would be disastrous if everyone were to lie whenever he wished to is completely undercut by the final clause of the same sentence, "which is, however it may seem, actually not the prevailing practice." For clearly if we are not sure that this is not the prevailing practice (and if we were sure, Singer would not need to mention it), we cannot be sure that the consequences would be disastrous if it were. Assuming that we do not now *always* lie whenever we wish, and that we wish to lie as often as Singer seems to think we do, one might argue that the long run consequences of adopting this practice would be desirable, for we would quickly learn of the bad consequences of frequent lying and thus soon would not wish to lie very frequently. Thus we would all become more truthful. This is in fact what occurs to many children. I do not claim that this argument is sound, but only that it is not obviously faulty. For that is enough to dis-

prove Singer's claim that lying is wrong because of what would happen if everyone lied. The fact is that we are much more confident that lying is wrong than we are that the consequences would be disastrous if everyone lied whenever he wished. Indeed, our willingness to agree that the consequences would be undesirable rests largely on our belief that lying is wrong *independent of its consequences*. For, if lying is wrong, then the consequences of everyone lying whenever he wished would be undesirable in terms of the general degeneration of moral character. But this is not, of course, a consequence one can cite in arguing that lying is wrong *because* of what would happen if everyone lied.

It is true that if everyone *always* lied, the consequences would be disastrous; for that would involve abandoning language and human culture. We may formulate an argument, from the principle of consequences: The consequences would be disastrous if everyone always lied, therefore it ought not to be the case that everyone always lies.⁹ Here, of course, the true and relevant response is "Not everyone will." I should say that there is *no* probability that everyone might always lie. For all of us, most of the time, our interests are best served by truth-telling; we have no inclination at all to lie all the time. Hence this argument, by itself, does not show that lying is wrong nor justify our limiting anyone's right to lie. If there were no other grounds for saying that lying is wrong, it would be in the same category as refusing to have children, or refusing to drink milk.

There are, of course, other grounds. But before turning to them, let us discuss briefly another fundamental moral rule which Singer says is derived from an application of GA. This is the rule that stealing is wrong. When he says that a moral rule is "established" by or "derived from" an application of GA, he seems to claim that the *logical basis* of the rule is GA; that the correctness of our identification of the act as "wrong" is dependent upon the use of GA; that, in his own words, the fact that if everyone did *x* the consequences would be undesirable is what "makes it wrong or constitutes its wrongness" (p. 120). But this cannot be the case with the rule that stealing is wrong. For, while it may be true that the consequences would be undesirable if everyone freely took things which belong to others, these consequences can occur only if there is a *prior* rule

⁹ Note that if we use this premiss in the generalization argument, the conclusion will be, "Therefore, no one ought *always* to lie." And this does *not* establish that lying is generally wrong, or that each act of lying is wrong—even if GA is valid.

identifying such action as wrong! For without such a rule there would be no concept of an object *belonging to someone*, and it would be impossible to perform the act in question. The concept (or institution) of property (or belonging to) logically implies the rule that stealing is wrong. Any community which has the institution of property must also have, in some form, the rule against stealing—it is the existence of the institution which “constitutes the wrongness” of stealing. In any community which does not have the institution, it is logically impossible for anyone to steal. Thus the rule, stealing is wrong, cannot, in this sense, be derived from an application of *GA*; for stealing must already be wrong before the acts described in the premiss of this application of *GA* can occur. A similar argument may, I think, be made about lying, and, incidentally, about the rule that one ought to obey the rules of his community.

But it may be contended that this is irrelevant to Singer's argument. For while it is true that the “impropriety” or “wrongness” of certain acts follows from the existence of some institutions, this is only a *formal* impropriety. The rules so established, such as that stealing is wrong, are so far only *formal* or *institutional* moral rules and not *fundamental* moral rules. And it is as fundamental moral rules that Singer is concerned with them. But though there is some point to this claim, we must insist that the formal or institutional moral rule is logically prior to any application of *GA*, and that the claim that the rule is “derived” from *GA* is at least misleading. For we do not begin with acts which are morally neutral in the sense that, prior to an application of *GA*, there is no reason at all for saying that they are wrong. Rather the existence of the institution creates the presumption that the act is wrong. It is only when this presumption is questioned that we must go beyond the institution for *further* reasons. And it is only when there are independent reasons for saying that *x* is wrong that the generalization premiss cannot be neutralized by the response, “But not everyone will do *x*.”

However, suppose we distinguish *formal* from *fundamental* moral rules and take Singer to be talking only about the latter. Does he show that these are derived from *GA*? We have already seen that when we apply *GA* to lying whenever one feels like it, the truth of the premiss of *GA* is much less

certain than the conclusion. Hence if there is some more plausible argument than *GA*, we would be foolish to rest our claim that lying is wrong on this dubious premiss. And even admitting only the principles acknowledged by Singer, there is a much better argument, not only for the rule against lying, but for most of the fundamental moral rules Singer mentions, i.e., against killing and stealing. For unlike acts which violate neutral norms or local rules, every individual act which violates a fundamental moral rule is such that its successful execution *assures* the occurrence of some undesirable consequences. For every successful act of lying, someone is deceived; and that is undesirable. For every successful act of stealing, someone loses something which belongs to him; and that is undesirable.¹⁰ Hence the wrongness of such actions follows immediately from what Singer calls “the true logical generalization of the principle of consequences,” i.e., if the consequences of each and every act of kind *x* would be undesirable, then each and every act of kind *x* is wrong.¹¹ Since these fundamental moral principles can be established by the principle of consequences, we may conclude that they do not depend on *GA*.

Thus when Singer's examples and arguments are analyzed, none of them shows that moral rules are established by the generalization argument. The generalization *premiss* is, indeed, evoked, but when it is effective it is the premiss of an application of *C*, not of *GA*. When it is taken to be a premiss of *GA*, the argument is usually invertible, or the premiss is incomprehensible, false, or insufficient to establish the desired conclusion. In fact Singer has not shown that *any* moral rule, of *any* kind, requires *GA* as its basis. And we have seen that some moral rules of each kind can be established without any reference to *GA*.

Adding these results to those of the other writers cited above, I conclude that Singer has not shown *GA* to be either valid or a fundamental moral principle. So far, we have no reason to accept any of his claims for its importance in moral reasoning.

VI

We may now return to the question raised at the end of our discussion of neutral norms. Perhaps the proper claim to be made is not that moral rules

¹⁰ This point is sustained, I believe, by the fact cited earlier in this paper, that about violations of these rules we are much more likely to ask, “How would you like it if someone did that to you?” than “What would happen if everyone did that?”

¹¹ A somewhat weaker argument, using what Singer calls the “generalized principle of consequences” also applies to these acts, i.e., if the consequences of doing *x* would be undesirable (in general or usually) then it is wrong (in general) to do *x*.

themselves are derived from *GA* in the senses we have discussed, but rather that, however moral rules come to be established, *our reason for obeying them* is to be found in *GA*. This seems perhaps a more plausible claim. If true, it would give *GA* some importance in our moral thought, though not the status Singer assigns it. But it is obviously not true either, if it means that we have *no* reason for obeying any moral rule, independent of *GA*. For the grounds on which moral rules are established also provide reasons for obeying them. I ought not kill because the consequences of my killing will be undesirable. I ought to obey the law about driving on the right side of the road, or paying taxes, because they are laws properly established in my community and therefore it is incumbent upon me to obey them. So *GA* does not provide our only reason for obeying moral laws.

Perhaps then the claim may be that all other reasons are insufficient; *GA* alone provides a sufficient reason for obeying moral laws. This claim is difficult to assess; for the sense of "sufficient" is not clear. If *logical sufficiency* (if we know what that means in this context) is at stake, then if we allow, with Singer, that *C* is a fundamental moral principle, it should provide a logically sufficient reason for obeying those rules which are established by its use, such as the rule that killing is wrong. In allowing that there are other moral principles at least as fundamental as *GA*, Singer has, in effect, already disallowed the claim that only *GA* can provide a logically sufficient reason for obeying a moral rule. If *psychological sufficiency* is our concern here, the case is no better; for if anything seems clear it is that no measures, not even the most drastic threat of punishment, are psychologically sufficient to persuade everyone to obey any law. We just cannot say of *any* consideration that only it will be psychologically sufficient. And it is quite unlikely that a person unmoved by the fact that a particular action *will have* undesirable consequences, would be moved by the very hypothetical argument that *if everyone did x*, the consequences *would be* undesirable—particularly when it is highly improbable that everyone will do it. We must thus reject any contention that only *GA* can provide a sufficient reason for obeying a moral law.

We have not found that *GA* is necessary for the establishment, defense, or normal functioning of

any moral rule. Must we then conclude that *GA* has no role at all in moral reasoning? No. Though its role is limited, there are contexts in which it has some persuasive force. These contexts are those in which there is an acknowledged law or rule, but a particular action falling under it is questioned because it is not at all obvious that the consequences of an isolated act of disobedience would negatively affect any specifiable individual nor that they would be undesirable either for the agent or on the whole. Such situations arise, for example, with regard to the requirements that everyone pay taxes or do military service. We may ask then: Why should I obey this law when it will be to my great disadvantage, and only of negligible advantage to anyone else? Here the response, if everyone were to avoid paying taxes the consequences would be disastrous, seems quite relevant. *GA* seems directly applicable; its conclusion fairly supported by the generalization premiss.

But even here, if we look more closely we may see that *GA* is but an unclear way of stating the real grounds for obedience. Such laws are generally enacted to achieve certain community purposes or goals: security, better education, efficient transportation, etc. The achievement of these purposes would generally be advantageous to every member of the community; as members of the community we ordinarily share the community purposes. General obedience to the law is usually a necessary condition for achieving these purposes. These are the factors which support the demand for obedience, which render irrelevant the fact that we cannot specify any individual who will be affected by our act, and which, perhaps, persuade us to obey. If we are committed to the achievement of the community purposes, and we recognize that general obedience to the law is a necessary condition of this, then we have a voice in the community demand that the law be obeyed. But then it would be unfair for us to evade the law ourselves while demanding general obedience to it. Hence, like others, we ought to obey the law. The ultimate basis of our obedience is our recognition of the purposes or values to be achieved. It is these factors which are obscurely brought to light by our use of *GA*. But only obscurely! The foundation of our morality is not in fact as negative as Singer's insistence on the importance of *GA* suggests.

IV. EMOTION AND THE CONCEPT OF BEHAVIOR: A DISPROOF OF PHILOSOPHICAL BEHAVIORISM

MORELAND PERKINS

IN defending what he calls philosophical behaviorism, exemplified by a proposition he thinks to be so self-evident it "asserts nothing"—that "to be angry is to behave in certain ways"—Paul Ziff has argued that a person's behavior may be quite as inscrutable to an observer as any emotion can be.¹ Despite Ziff's claim that it is no move at all, this is a bold one. Instead of demolition, or destruction by attrition, the enemy's battered stronghold is occupied. But after recovering from the first shock of this thrust, the opponent of behaviorism must be heartened; by the very clarity of Ziff's advance the inadequacy of the behaviorist's resources is made more distinctly visible. For one's feelings, like one's thoughts, can be private; there can be no sane doubt of this. But when we speak of private behavior, a moment's reflection will incline us to believe that the sense of "private" we employ has changed. There is the public behavior of a public figure and his private behavior. But shall we say this: that the privacy of some of the public man's thoughts and the privacy in which some of his emotions are felt is the same domain as the one wherein his private behavior lies? One is inclined to say no. And I think indeed it is not. The privacy of some thought and emotion is a solitude that makes the realm in which *all* behavior must lie, by contrast, a public domain.

To understand the concept of emotion one must know that emotion can sometimes be concealed from those in the presence of whom it occurs, even when it is directed toward them. But behavior cannot, in similar circumstances and from persons similarly related to it, be *concealed*: the concept of behavior does not allow it.

Behavior must be public. Emotion may, but need not be. The concepts are ill-suited for equation; emotion cannot, in general, be simply a way of behaving.

However, the sense of "public" in which it is true that behavior must be public may not, in the face of Ziff's arguments, define itself during a moment's reflection. Indeed there is perhaps no one sense of the word "public" that comprehends the several dimensions of behavior I wish to suggest by using the word. And I shall *both divide and stretch* the meaning of "public" in the course of defending my contention that, unlike emotion, behavior must be public. But the verbal point aside, what matters is this: although all that is required of the most ordinary sort of behavior is not required of all behavior, in each instance of behavior something is demanded that is not required, in comparable circumstances, of emotion. I find it suggestive to refer to each of these somewhat different criteria of behavior as a kind of *public* character. Others might prefer another word, or prefer to avoid attempting to gather together under a single rubric the differing tests to which, from situation to situation, action is submitted before it is labeled "behavior." I shall be content if I can show how the concept of behavior contains, in the form of case by case prescriptions for the use of the word "behave," something that so distinguishes it from the concept of emotion that it cannot serve in the analysis of emotion as the philosophical behaviorist thinks it will.

Ziff admits that "You cannot in fact always find out whether or not I am angry. I may be artful at concealing my anger and I may refuse to tell you."² But in opposition to the critic of behaviorism who maintains that nothing of the sort can be said of a man's behavior, Ziff continues, "Neither can you in fact always find out whether or not I am flexing my abdominal muscles. I will not tell you and no one else can."³

We can accept Ziff's claim that another person may be unable in fact to tell whether or not Ziff

¹ In "About Behaviorism," *Analysis*, vol. 18 (1957-58), pp. 132-136; reprinted in V. C. Chappell (ed.), *The Philosophy of Mind* (Englewood Cliffs, N.J., Prentice-Hall, 1962).

² *Loc. cit.*

³ *Ibid.*

is flexing his stomach muscles. But to characterize something a person does as in this way unknowable in fact for others and then to say that it is a way of behaving is a mistake. Flexing one's stomach muscles, under this description, is no more a way of behaving than the unfriendly thoughts a person may have as he smiles and to all appearance acts amiably comprise a way in which he behaves. The person's behavior was friendly; his thoughts were not. We can imagine suggesting to someone that he covertly flex his stomach muscles, as part of our advice about how to sustain some piece of behavior. For example, when the behavior desired is standing motionless and at attention for an hour, it may be that a man can help himself by flexing his stomach muscles, if the flexing doesn't show. He might also be helped in that situation by thinking of what it would be like to have to stand that way for ten hours instead of only one. But neither the thinking nor the hidden flexing is itself a way of behaving. Flexing your muscles in cadence while participating in a class in calisthenics is behaving, but then the flexing is both public and social. Normally, concealed flexing can at best be a way of *trying* to behave, as when by flexing my muscles I try to maintain my military stance.

The man who is sustaining the difficult posture may conceal from the inspecting officer his thoughts, his emotion, and his abdominal tricks; his behavior it is not possible for him to conceal. His anger may be hidden; his behavior may not be. In this situation, what he hides from the world in the way of bodily movement ceases to be behavior. But that he should hide what he feels does not cause it any the less to be an emotion, say anger, that he continues to feel. Here the rules governing the use of the words "behavior," "emotion," and "anger," make it impossible that certain hidden emotions should be ways of behaving. So an emotion cannot always be a way of behaving.

Typically, how a man behaves is how he comports himself in the presence of other persons. Normally, to speak of a man's behaving in a certain way is to call attention to a portion of what the man does in the way of relating himself to others that is experienced by others at the time it happens as the sort of thing it is now said to be. In this sense of the word "public," behavior is normally public.

The philosophical behaviorist may reply that normally emotion, too, is public. But this normal publicity, if it exists, is not, I think, a conceptual

norm in the sense of a standard or mark that emotion has always to meet. Whereas the publicity of behavior is a norm, a standard which the concept of behavior requires that action meet if it would qualify as behavior. Publicity is always required of behavior; it is not always required of emotion.

When behavior is not public in the *simple* way just explained, it must meet somewhat different but connected standards. The best way to see that this is so is, I think, to examine what happens when deviation from the norm of straightforward publicity is permitted to behavior.

Although concurrent appreciation by others provides the standard context for a piece of human behavior, behavior is not always appreciated at the time it occurs for what it is. We speak of how a soldier behaves in the presence of the enemy, or under fire (of the enemy). But we also say that a soldier who is alone and wounded in difficult but deserted terrain and who perseveres, say, in carrying a message to its destination behaves with courage; then much of what he does during the period when no one is present to observe him is behavior. However, it is evident that we are here viewing the man's action as social, in a definite and common sense of the term. And this is essential to our here speaking of his *behavior*. We count him as a member of a group whose fortunes depend in some measure upon his actions, and we think of him as acting in cognizance of this dependence of others upon what he does. His is the sort of action whose origin and natural completion is public in the ordinary sense and social in a clear one. We count what he does as behavior because we conceive him to be acting as a member of a group to which he is bound by recognized conventions in ways that define certain sorts of action under some conditions as obligatory, expected or required, or something of the sort. But no condition of this kind is necessary in order that this unobserved and isolated man should be afraid, awed, or elated. Independently of his being a soldier, or anything like one, the isolated, unobserved man who happens to be a soldier may be elated, awed, afraid. It would appear, then, that at least some instances of our isolated soldier's experiencing an emotion cannot be ways in which he behaves.

But of course this appearance could turn out to be deceptive—possessing a social character might not be a *necessary* condition of "solitary behavior." To this scruple we shall return near the end of our discussion. But it is necessary first to undertake a

- more thorough examination of the possibilities open to behavior and to emotion when each occurs in the presence of observers.

Ziff claims that a person's behavior can be inscrutable in the presence of persons normally equipped to observe him, and inscrutable in the way that a person's emotion is inscrutable when it is concealed. An emotion is simply a way of behaving, according to Ziff. And what is distinctive about Ziff's essay is the subtlety with which he argues that behavior may be every bit as impenetrable to an observer as concealed emotion may be.

Now there is a kind of situation in which a person can be said to be behaving and yet to be doing something *like* concealing his behavior from persons present and observing him. But it is a special sort of situation. And what happens is not exactly that behavior is concealed. The behavior of a spy in the performance of his duties provides one sort of paradigm.

Let the spy be a woman. She may be using great ingenuity in softening up an enemy officer and "leading him on"; she may be guiding the conversation into areas useful for her hidden purposes; she may behave calmly in a moment of danger. The enemy officer in whose presence all this goes on may be unable in fact to discover that she is behaving in any of these ways. It would not, even here, be correct to say that the woman is "artful at concealing" her behavior. She does not *hide* her behavior. She hides from him, rather, the meaning, the significance, of that behavior of hers which he does observe. The effect of this concealment is that by means of the behavior which the officer does observe, the woman who is a spy behaves in ways that he does not appreciate for what they are. Now, what makes the part of the spy's action that the enemy officer cannot at the time appreciate, *behavior*? I think the answer is the same, essentially, as the answer given to the same question concerning the unobserved and isolated soldier: we call the spy's action behavior because we are viewing it as a piece of teamwork, as I will now call it. The woman is operating as part of an espionage outfit, a "team," and the fact that her action is a piece of teamwork justifies us in calling it behavior, in this case. In *this* sense of "public" her action, like the unobserved soldier's, is behavior because it is public. On the other hand, whatever the spy's

emotions may be, the fact that she is engaged in teamwork—though of course it may affect or even determine her emotion—does not enter into our *justification* for characterizing what is happening to her as a case of her feeling an *emotion*. Consequently she may feel toward the enemy officer, but hide from him, an emotion that has nothing to do with her being a spy or anything like a spy.

But the wary reader may suspect that I have chosen my examples too carefully. Surely, he will remark, behavior may occur in the presence of others properly placed and equipped to observe all of one's movements and yet neither be appreciated by those others for what it is nor be a piece of collaboration with someone absent. So it may.⁴ But before we look head-on at this hard fact, let us first consider a transitional case that illustrates the other side of the coin of teamwork.

Suppose that George goes to the High Life Bar with John. In doing this George may be behaving in a certain way toward Georgianna, for he has solemnly promised Georgianna that he would not go to the High Life with John again. George's behavior toward Georgianna is not a case of teaming up with John when John knows nothing of George's promise to Georgianna; and George is certainly not then teamed up with Georgianna. It may be that John cannot in fact know at the time that all night long George has been behaving in a certain way toward Georgianna, although John is present, attentive and normally equipped for observing George. George won't tell John, and no one else then and there can. However, in one important respect, this case resembles the case both of the spy and of the isolated soldier: what George does in the High Life Bar that John is unable to appreciate is properly characterized as a way of *behaving* only in so far as it is proper to think of George as acting in his role of (delinquent) partner with Georgianna in something like a contract, here a promise. George and Georgianna form a group of the kind mentioned in connection with the isolated soldier, and George is still acting as a member of it even as he repudiates, in a measure, his membership in it. At least it is essential that we so view George when we count what he is doing that John must miss as *behavior*. Again, no condition of this kind limits what may properly be viewed as an emotion which George feels while at the High Life Bar and that John cannot know he feels: he may

⁴ I am indebted to my colleague, Lawrence Resnick, and to Mr. Peter Winch, for insisting—and, as will shortly appear, to Jane Austen for convincing me—that it may.

feel what fails to pass this test for behavior.⁵

Before we turn to the hard case of behavior which so far resembles hidden emotion that it takes place in the presence of the person to whom it is directed, it is inscrutable, and yet it is neither a bit of teamwork nor a case of betrayal, it will also be useful to remind ourselves of an *extended* sense of "behave" that is current among scientists, and of the reasons why the philosophical behaviorist of Ziff's sect must eschew this extended sense of the word.

Atoms and molecules are said to behave in certain characteristic ways. Sticks, stones, and bones may all behave, according to an extended use of the word "behave." Behaviorism with respect to emotion—the position that asserts of each emotion a proposition like "to be angry is to behave in certain ways"—might be understood as employing "behave" in this extended way. I imagine that some scientific behaviorists have used the word with this wider extension. However, this option is not open to the "philosophical behaviorist." In his essay Ziff speaks against the propriety, for his purposes, of this extended use of "behave." And his care in this matter constitutes the crux of his defense of philosophical behaviorism, specifically of his claim that behavior can be (and is) sometimes just as inscrutable when it occurs as emotion sometimes is. Ziff writes that "it does not follow that if certain teeth and jaws are moving in certain ways [which is in principle always knowable by others at the time] then a certain organism is behaving in certain ways."⁶ The conditions necessary, according to his account, for it to be the case that *I am gnashing my teeth* (behaving) are different from, in that they are additional to, those that are necessary for it merely to be the case that my teeth are gnashing. Mere gnashing of my teeth could be induced by direct stimulation of the relevant muscles by another person; this is not behaving. Behaving is something *I do*, that is to say, something *I* do. Although it is always possible in principle for someone else to know at time *t* (without my telling

him) what movements and postures characterize my body at time *t*, it is not always possible in principle for someone else to know at time *t* (without my telling him) that I am behaving in a certain manner at *t* when I am. Behavior, in Ziff's account, is never mere bodily movement, orientation, or change; it is, rather, what I accomplish by their means. However fully scrutinizable the situation, movements, and other changes of my body's state are, what I am up to by means of them is sometimes quite inscrutable to an observer. And according to Ziff it is what I am up to that constitutes my behavior. But the extended use of "behave" countenances no distinctions of this sort. Movement, posture, bodily changes merely as such suffice for behavior, according to the extended sense of the word: teeth can be said to behave, according to this extended usage, as can jaws, and the whole organism can be characterized as behaving when its teeth are made to gnash by means of artificial stimulation of the muscles, or when its heart beats.

Now, barring a "vitalistic" biology, a behaviorist who employs this extended sense of "behave" is simply a materialist. And it is not materialism that Ziff is defending. When Ziff disaffiliates himself from the notion of behavior as simply change of bodily state, he is appealing to the concept of human behavior that plays a part in the affairs of ordinary life. And my tack here has been to quite agree with Ziff that in everyday life the criteria of *behavior* are richer than those for bodily movement, orientation, and change of state, but to insist that there are more criteria than even Ziff has bargained for. Our practical concept of behavior encloses a narrower space than Ziff has realized. It not only excludes what Ziff has so perceptively discerned that it does exclude—mere bodily movement, bodily orientation, and the like; it excludes too what Ziff thinks it can encompass—whatever is private in the way emotions are private when they are most so.

Now let us return to the comparison of behavior's occasional inscrutability with that of emotion. Ziff

⁵ There is, too, a trivial case in which, even without the complications of commitment to or conspiracy with absent persons, it is possible for behavior to be unobserved (although not either concealed nor hidden) in the presence of persons normally equipped and as well placed for observation as it is possible to be. In my presence a person may insult me or compliment me and I may fail to realize it simply because I am not paying proper attention. A person may wave to me without my noticing it, although I should have noticed if I had been looking at him or if I had not been distracted at the moment. But what is evident about behavior of this kind is that ordinarily it is behavior only insofar as the person at whom it is directed and who fails to observe it *would* have appreciated it if he had simply bothered to pay attention. To characterize these unnoticed actions as ways of behaving is, ordinarily, to imply that they meet this condition. However, nothing of this sort is true of emotions that go unnoticed in comparable circumstances: the test for emotion may be passed when the test for behavior is failed; not every emotion will qualify as behavior.

⁶ Ziff, *op. cit.*

claims that the concept of behavior that figures in everyday affairs embraces behavior that is inscrutable at the time by observers who are as well placed as observers may be. I have admitted this possibility, but suggested that it can be realized only if special criteria are met which do *not* bind emotion that is inscrutable under comparable conditions. The chief of these tests that has emerged is that in some circumstances the inscrutable action qualifies as behavior only in virtue of its being a piece of teamwork—or betrayal—and in that sense by being public. But it will be necessary to consider a case of inscrutable behavior that is neither a piece of teamwork nor a case of betrayal. And I shall do this by discussing a portion of what the novelist Jane Austen, in *Sense and Sensibility*, wrote concerning the behavior of a young man named Willoughby toward a young woman named Marianne Dashwood, that turned out to have been inscrutable to both Marianne Dashwood and her family, but not a piece of teamwork.

First, however, I will again digress and discuss, from the same novel, something simpler and more closely analogous to Ziff's hidden gnashing of the teeth.

In the passage that follows, Elinor Dashwood ("she") is comparing her own estimate of Colonel Brandon's ("his") romantic interests with the estimate made by her older friend, Mrs. Jennings: the latter thinks that Elinor herself is the object of these interests; Elinor (correctly) believes that her younger sister, Marianne Dashwood, is.

... she could not help believing herself the nicest observer of the two: she watched his eyes, while Mrs. Jennings thought only of his behavior; and while his looks of anxious solicitude on Marianne's feeling, in her head and throat, the beginning of a heavy cold, because unexpressed by words, entirely escaped the latter lady's observation, —she could discover in them the quick feelings and the needless alarm of a lover.⁷

In the situation being discussed the expression of Brandon's that Elinor perceived was something that showed itself on a countenance not occupying the attention of anyone but Elinor: we know that Brandon would be sitting in a corner, so to say, out of the public eye. It is evident that Jane Austen denies that Brandon's emotion here takes the form of *behavior*. The look expressive of Brandon's feeling toward Marianne is no part of how he is *behaving* on this occasion. For he is not, when it occurs, en-

gaged in speaking to Marianne or to anyone about Marianne, nor in offering his services to Marianne in any way; there is nothing he is up to in the way of behavior into which these expressive looks of his enter as a component. His emotion is, though not quite concealed, not quite public, either; so it is not, according to a discrimination Jane Austen herself explicitly makes, behavior.

This example of Colonel Brandon's non-behavioral activity in Marianne's presence proves that we cannot say that actual appreciation by another person will in *every* context suffice to make what could (we suspect) *become* a piece of behavior become one: the character of the larger context into which both the activity and its appreciation enter can be crucial in determining whether or not the activity is public in the way it needs to be in order to be behavior. How this works will be made more evident if we compare this account with Jane Austen's account of another situation that included Brandon and Marianne.

We now encounter Brandon *meeting* Marianne for the first time after the illness of hers, that he feared, has come and has passed its crisis. This time the mother of the two Dashwood sisters is present and she is the observer with whom observer-Elinor compares her own perceptions. And this time the sort of reaction in Brandon that was not behavior before is now *behavior*:

His emotion in entering the room, in seeing her altered looks, and in receiving the pale hand which she immediately held out to him, was such as, in Elinor's conjecture, must arise from something more than his affection for Marianne, or the consciousness of its being known to others; and she soon discovered, in his melancholy eye and varying complexion as he looked at her sister, the probable recurrence of many past scenes of misery to his mind. . . .

Mrs. Dashwood, not less watchful of what passed than her daughter, but with a mind very differently influenced, and therefore watching to very different effect, saw nothing in the Colonel's behavior but what arose from the most simple and self-evident sensations. . . .⁸

Mrs. Dashwood sees in Brandon's "looks" and in his changes of complexion, in *this* situation acknowledged by the author to be *behavior*, Brandon's feeling for Marianne. Elinor sees this in it too, though she sees more as well. (But the more we can ignore.) Certainly Ziff would not be obviously wrong to say that in this second scene

⁷ Ch. 42.

⁸ Ch. 46.

Jane Austen does count Brandon's behavior as (at least part of) his emotion: I think she does. And the behavior has for its vehicle the very sort of thing that showed itself in the earlier scene only to Elinor: the expression of Brandon's eyes and (what was not mentioned in the earlier scene but could have been) his "varying complexion." Here the look of the eye is behavior; there it was not. In the first situation Brandon's emotion was *not* behavior. Here it is, at least in part. Why is this? Because the expression of his eyes now enters into Brandon's act of greeting Marianne, when he is the chief actor in the scene. In the earlier scene it was scarcely necessary for Brandon to conceal his emotion—he himself was too obscure a component of the situation for his reactions to Marianne to be anything more than a private affair that happened to be noticed by Elinor. Now he has scarcely the power to keep his emotion from *becoming* behavior; and certainly he does not try. According to Jane Austen, before, his emotion was one thing, his behavior another; now, the two coincide. And the difference between the two cases is a difference between Brandon's facial expression's possessing a certain kind of public character and its failing to possess it.

But now let us face the problematic case of deceit on a large scale, the case of young Willoughby. Was his behavior as hidden as emotion may be?

At Barton, where the Dashwoods were staying, Willoughby behaved in a way that led all the Dashwoods to believe he was in love with Marianne Dashwood and would propose marriage to her. And Marianne fell in love with Willoughby. Then, suddenly, Willoughby took his leave. Later he wrote Marianne of his now impending marriage to another woman (to whom he had not been engaged while at Barton). And he married the other woman. Then Marianne's distress developed into a serious illness.

In deceiving Marianne during the early days of their acquaintance Willoughby was, he later confessed to Elinor, pursuing a purely private end, his own pleasure, nothing more. Marianne (and her family) did not then appreciate his behavior for what it was. But it was behavior. Listen to Elinor, speaking to Marianne long afterwards:

"The whole of his behavior," replied Elinor, "from the beginning to the end of the affair, has been grounded on selfishness. It was selfishness which first made

him sport with your affections; which afterwards, when his own were engaged [to you], made him delay the confession of it, and which finally carried him from Barton. His own enjoyment, or his own ease, was, in every particular, his ruling principle.⁹

Three instances of Willoughby's behavior are here singled out. It is only the first that offers a challenge to my contention that behavior, unlike emotion, must always be public. His departure from Barton was a public performance. And his failure to propose to Marianne once he *had* fallen in love with her, was behavior in so far as it was the knowing *failure* to perform an act that must, if performed, have been appreciated at the time by Marianne, in order to have been a confession of love. But the first instance of behavior cited, Willoughby's "sporting with [Marianne's] affections," deviates from the norm of publicity in almost exactly the way that our spy's behavior did, with this critical difference: Willoughby's sporting with Marianne's affections was not a piece of teamwork, and consequently it was not even in that way a public performance. The obstacle to verification of my claim that, unlike emotion, behavior must be public, resides in the fact that not only is Willoughby's "gallantry" behavior, so is his sporting with Marianne's affections: what Willoughby does that *no one* but he can appreciate at the time is also behavior. Yet Willoughby has no absent colleague or partner with whom he is collaborating or whom he is betraying in this behavior.

Of course it is still incorrect to say that Willoughby was concealing his behavior from Marianne. On the contrary, what he was concealing were his feelings, his thoughts, and his intentions. When, after his marriage and during Marianne's illness, Willoughby confesses to Elinor Dashwood, he confesses to what we might call a certain state of mind; and he thereby admits to having behaved quite as Elinor now says to Marianne that he did behave.¹⁰ He does not *confess* his behavior; he cannot do that to one who was on the scene. That part of Willoughby's behavior that is characterized as his sporting with Marianne's affections, consists, to put it crudely, in his "gallantry" (itself behavior) together with the intention with which he was gallant, or the motive of it. And the intention was hidden. Indeed his intent was concealed, with the result that his behavior was inscrutable. His feelings

⁹ Ch. 47.

¹⁰ See ch. 44.

too were concealed, with the same result. But his behavior was not concealed from the Dashwoods. To say so would be to speak nonsense.

The problem is this: what is it about Willoughby's sporting with Marianne's affections that makes it behavior? Must this deviation from behavior's norm of concurrent appreciation by another also, as in the case of the isolated soldier, of the spy, and of George in the High Life Bar, meet another, related standard in order to be behavior, a standard, too, that emotion need not meet? I think that it must.

First, when Elinor views Willoughby's sporting with Marianne's affections as behavior, she views what he did as meeting or failing to meet certain standards regarding how one person should act toward another person. To classify as *behavior* something that cannot at the time and by those in whose presence it occurs be appreciated for what it is, something too, that can be understood only by a reference to concealed and purely personal ends, is to mark that something out as subject to appraisal in the light of standards specifically applicable to how one person *acts* in relation to another person.¹¹ But *nothing exactly like this applies to emotion*. It is perfectly possible to say of something which Willoughby in those same circumstances experienced that it was an emotion, when it is impossible to apply to it any standards indicating how people should *act* in relation to one another. Hence Willoughby might experience an emotion that *could* not be behavior, even by the criterion of behavior that permits behavior to be both inscrutable and non-conspiratory.

Second, to view as behavior what Willoughby was doing that Marianne could not at the time appreciate, it is necessary either that we think of it as having had an effect upon Marianne (as in this case we know that in the end it did) or, in case it did not have an effect or we do not know that it did, that we think of it in terms of the kind of effect upon Marianne it might and normally could be expected to have. Willoughby's inscrutable sporting is behavior *because* it is (at least potentially) a way of acting *upon* another person. Now we can see the *necessity* there is, according to the concept of behavior, for some component of behavior to be bodily motion or posture, or some other outwardly discernible bodily change: with-

out one of these even the potentiality of an effect upon another person is, to modern common sense, lost. To say no more than that Willoughby behaved is already to *say* that he did something that had or could have had an effect upon another person (and so was at least to that extent public). To say Willoughby thought something to himself is also to say that he did something; but it is not, in itself, to commit ourselves to even so much as the possibility that what he did should have an effect upon another person. And to say that Willoughby felt an emotion—or to say that he was angry—is not even to say that he *did* anything! Even less, then, is it to imply that he did something having a real or potential effect upon another person. In feeling an emotion a man commonly suffers effects, we acknowledge; but whether he also affects the other person present depends, for one thing, upon whether or not his emotion shows itself forth—which it need not do. So far as the implications of our words go, it is not *necessary* that even so much as a symptom of the emotion should be manifest. Sometimes—when for example it is effectively concealed—our emotion is permitted to be wholly out of sight. But this our behavior cannot be, for it *must* be capable of affecting the other person.

The lesson learned from Willoughby can now be applied retrospectively to our discussion of the two different situations in which Colonel Brandon reacted to Marianne Dashwood's presence with a change of facial expression. In regard to the one situation we discovered that, according to the author, Colonel Brandon did not behave, in so far as he looked in a certain way at Marianne, and in regard to the other situation that he did thereby behave; and so in the one case Jane Austen viewed Brandon's *emotion* as consisting, at least in part, of behavior, and in the other she viewed it as not involving behavior at all. We can see now that *one* mark of the expressive "look" that did count as behavior was this: what Brandon then did when he looked in a certain way at Marianne was capable of having, and would *naturally* have some *effect* upon Marianne; because it was a significant part of how he entered the room and greeted her when all eyes, including Marianne's, *would* be upon him. Whereas when Brandon's "looks" were not part of his behavior, in that first scene before Marianne had become seriously ill, we understand

¹¹ In generalizing, I say "acting in relation to," not "toward." The *simple* paradigm for acting-in-relation-to-a-person, as distinguished from acting-toward him, is provided by the relation in which the action of person *A* stands to person *B* when *A* does not act *toward B*, *A* does act, the situation in which *A* acts includes *B*, and anyone situated as *B* is would naturally be expected to *notice A's* action.

that how Colonel Brandon, in his obscurity, looked at Marianne was, as Brandon himself well knew, something quite outside the range of occurrences capable of having an effect upon Marianne.¹²

It was, however, the sort of behavior illustrated by Willoughby, which comes so close to being concealed from actual observers in the way that emotion sometimes is, that offered the most interesting obstacle to the proof of my contention that in each of its instances behavior needs some character that emotion, in comparable circumstances, may lack. And the study of Jane Austen's account of Willoughby and the Dashwoods has, I think, provided good reason for drawing the following conclusion. Inscrutable behavior is always *action*. Furthermore, it is action that has, at least potentially, an effect upon another person. So even inscrutable behavior must always be at least in part discernible to another person at the time it occurs. An emotion experienced in another's presence, on the other hand, (1) need not be an action, (2) need not even potentially have, in itself, an effect upon another person, and (3) need not

show itself forth even in part. And here, to say that emotion need not meet any of these criteria is to say that in practice we do, often enough, count as emotion what fails to meet any of these tests—in short, that some emotions do not meet them. We have, then, a threefold ground for the proposition that some hidden emotions cannot be the sort of behavior they most nearly resemble, the sort exemplified by Willoughby's inscrutable sporting with Marianne's affections.

I conclude that philosophical behaviorism is, as yet, a false doctrine. Perhaps it is not now possible to know that it must always be false. Or perhaps it is: for if the implications of my argument are pursued, I think it will appear that this metaphysical thesis could become true only if it should come to pass that men and women *only* behaved. But the only sense of "behave" in which it is logically possible for something *only* to behave is the extended sense in which a stone "behaves" when it falls. So if *we* only behaved we should be as stones; and this is a logical impossibility.¹³

*State University of New York
College at Cortland*

¹² With the lesson learned from Willoughby I think that we can also meet the objection made to my treatment of the isolated soldier, to the effect that the social condition I there singled out is not *necessary* to all "solitary behavior." Alone in his room and in collaboration with or betrayal of no one, it may be said, a man could engage in calisthenics. And in case it came to be known by others that he had exercised in this way, his activity might be said to have been a way in which the solitary man behaved. I think that such activity would not ordinarily, and spontaneously, be classified as "behavior." Yet certainly it could be; and it might be rather a barren research to undertake to determine exactly how far the ordinary concept of a way of behaving is extended beyond its normal range when it is made to apply to such a case. But this much is now evident: we should call such goings-on *behavior* in virtue of their being actions that would show themselves at least in part to any unobserved observer imagined present. As behavior they are things the solitary man does with his body, and this is necessary; necessarily, therefore, they are "public" at least in part; they are, necessarily, at least in part discernible to any ordinarily equipped observer imagined on the scene. But sitting alone a man may feel, as he may think, what no one translated to the spot could, even in part and in principle, by ordinary means observe. The solitary man's emotion may be wholly hidden, and then it is not, even in this somewhat extended sense, behavior.

¹³ Of course instead of imagining all action losing its intent and becoming, so to say, mere movement, one can perhaps imagine (or imagine that one imagines) all intent, thought, and feeling becoming public. We humans should perhaps then have become a sort of community of convivial witches. But there would then be no need for the concept "behavior," since *everything* would be public. Nor, I suppose, should we any longer be we.

V. RULES, DEFINITIONS, AND THE NATURALISTIC FALLACY

G. P. BAKER AND P. M. HACKER

IN *Principia Ethica* G. E. Moore introduced the term "the naturalistic fallacy" to name a particular mode of erroneous argument in moral philosophy. The specific error which he attacked was the attempt to define "good" in terms of what he called natural or non-natural predicates; Moore thought that this was impossible. This fallacy, which might more appropriately be called the "definitional fallacy," is a special case of a wider mistake. Any attempt to derive norms or values from "descriptive matters of fact" or definitions has been termed "descriptivism" (Hare) or the "naturalistic fallacy" proper (Prior and Hare). Both the fallacy and its demonstration by so-called "non-naturalists" have a history¹ that stretches back in English philosophy long before the time of Moore. Indeed Moore's own teacher at Cambridge, H. Sidgwick, pointed out both the value and the norm aspects of the fallacy. It is, however, primarily since Moore that philosophers have been preoccupied with this issue. Partly because of Moore's influence, attention has been concentrated upon only one aspect of the error, viz., the definitional fallacy with regard to value words. Although there has no doubt been considerable discussion of the alleged logical "gap" between norms and "statements of fact," this problem has not hitherto been adequately treated.

This inadequacy is evident in the prevalence of a certain kind of proposed refutation of the naturalistic fallacy and in the inconclusiveness of the criticisms of these arguments.² The kind of argument in question could be schematized in the following way. A "self-evident principle" or "ana-

lytic truth," e.g., "One ought to do one's duty" or "Promises ought to be kept," is combined with a "statement of fact" identifying a particular case as being, e.g., *A*'s duty or *A*'s promise, to yield syllogistically the conclusion that, e.g., *A* ought to do a particular act ("his duty"). The conclusion is taken to be "prescriptive" or "evaluative" or a "value judgment." Since the conclusion is interpreted as a "full-blooded" "ought"-statement, and since the premisses are allegedly no more than empirical statements and definitions, such a derivation, if valid, might be held to constitute a counter-example to the non-naturalist thesis.

In this article we shall discuss a particularly persuasive form of this argument, to be called "The Chess Derivation." To recognize something as a king in chess, it might be claimed, is to grant that if it is in prise (i.e., threatened or exposed) it must be moved out of prise.³ This, it might further be argued, is true by definition: it is part of the meaning of the term "king" in chess that if in check the king must be moved. So if the king is in check or in prise and under such circumstances one of the players utters the word "Check!" then in virtue of the meaning of the word "king", it would follow that the king must be moved. But how is it possible that definitions and descriptions of word-usage (e.g., of the performative "Check!") can commit us to the view that something must be done?

In our criticism of this type of argument we shall concentrate in turn on each of the two premisses and on the conclusion. First, we examine that status of the major premiss which is alleged to be

¹ See A. N. Prior, *Logic and the Basis of Ethics* (Oxford, 1961).

² See particularly J. R. Searle, "How to Derive an 'Ought' from an 'Is'," *The Philosophical Review*, vol. 73, (1964), pp. 43-58; Max Black, "The Gap Between 'Is' and 'Should'," *The Philosophical Review*, vol. 73, (1964), pp. 165-181; R. M. Hare, *The Language of Morals* (Oxford, 1961), pp. 43-44; R. M. Hare, "The Promising Game," *Revue internationale de Philosophie*, vol. 18 (1964), pp. 398-412; M. F. Cohen, "'Is' and 'Should': an Unbridged Gap," *The Philosophical Review*, vol. 74 (1965), pp. 220-228; and James and Judith Thomson, "How Not to Derive 'Ought' from 'Is'," *The Philosophical Review*, vol. 73 (1964), pp. 512-516.

³ Or, of course, the threat removed or blocked. For the sake of brevity we shall exclude such possibilities from further consideration. Similarly, in discussing the rules for the movements of the king we exclude from consideration the possibility of castling.

analytic or true by definition. Here we discuss the relations between rules and definitions in a rule-determined practice or "institution." Subsequently we examine the case for identifying the minor premiss as a mere "statement of fact." Finally we consider how to characterize the force of the conclusion.

* * *

The first limb of our argument is an examination of the major premiss. It is obvious that not all rules which may be used as the major premisses in syllogistic normative arguments can be construed as analytic or definitional truths (e.g., the rule of etiquette that in eating one must hold the knife in the right hand is surely not a definition of "knife" or "right hand"). On the other hand, that promises ought to be kept might be considered to be part of the meaning of "promise." Thus the argument we are examining is especially concerned with rules of the latter type, sometimes termed "constitutive rules."⁴

Numerous human activities, e.g., playing games, creating and dissolving legal relations as in marriage and contract, and certain forms of political activity, are only possible in virtue of rules. Such activities have been termed "practices."⁵ A practice is defined, determined, and regulated by what we shall call its "constitution." It is necessary that the constitution of a practice consist of both rules and definitions (see below). It should be stressed that the rules in question are not meaning rules, but norms or rules guiding and determining behavior, specifying what is prohibited, obligatory, or permitted within the practice. The relationship between rules and definitions may be quite complex. Thus rules may appear in definitions; the defined symbols must appear in the rules; and the definition of a practice may itself include or refer to the rules which determine it. These three points are illustrated by logical systems, where the rules of inference are equally essential to the specification of the system as the definitions, since they determine what "moves" are permissible in deductions. But as logical systems raise special difficulties, we shall illustrate the complex relationship of rules and definitions with the example of an interpreted constitution of the practice of playing chess.

The terms in the formulation of the rules of a constitution fall into three logically distinct categories: variables, constants, and modal operators. This may be illustrated from a possible formal constitution for chess. Let there be two variables, one taking squares as values, the other taking pieces. Let the squares be designated by the symbols '*A1*', '*E2*', etc., and the pieces by the symbols '*K*', '*Q*', '*R*', etc. The constants in this system will be of two types: logical constants and the non-logical constants characteristic of this constitution (chess-constants); among the latter will be such primitives as "standing on," "capture," and "move." The formation rules will specify the grammar of statements built from this vocabulary, e.g., that pieces stand on squares (and not: pieces on pieces). The symbols '*K*', '*Q*', '*R*', etc., can be introduced by formal definition in terms of the variable "piece," the symbols designating the squares, and the constants "stand on" and "move." The rules of chess, e.g., "If the king is in prise, it must be moved," will appear as the transformation rules of this system. Now, this schematic constitution would, if fully developed, be adequate for playing mental chess or for operating a written chess calculus. However, in as much as one wants to play chess on a board in the normal way, this uninterpreted calculus is as yet useless. What must be done is to correlate the terms of the calculus with elements and situations in the world by defining what is to count as a "piece," a "square," "a piece standing on a square," etc., and so giving an interpretation.⁶ One must of course distinguish two senses of "definition": giving interpretations to the primitive symbols, and giving formal definitions in terms of the primitive symbols. The fewer the primitives, the fewer the interpretations necessary, but the more numerous the definitions required for convenient specification of the constituents of the game.

Having sketched in this background, we can now move on to demonstrate one fallacy in the Chess Derivation. Let us consider in what sense it is possible that it is the meaning or part of the meaning of "king" that if it is in prise the king must be moved.

Is it conceivable that the king be exhaustively defined as the piece which must be moved if in prise? If this were the case, we should as yet have

⁴ Searle, *op. cit.*, p. 55.

⁵ J. Rawls, "Two Concepts of Rules," *The Philosophical Review*, vol. 64 (1955), pp. 3-32.

⁶ Of course the constitution of chess in a sense requires an interpretation even for mental chess or a written calculus, since the type-symbols in the rules must be correlated with the type-symbols in the game.

same form would apply to each of the other elements.

Characteristic of many games, including chess, is the utterance of performatives as an integral part of the game, e.g., "Check!" in chess, "Out!" in baseball, etc. The performance of such speech-acts, not to mention such extraludistic performances as promising and naming, are, as J. L. Austin has pointed out, subject to various forms of "unhappiness" or "infelicity." The happiness conditions of a speech-act are the truth-conditions¹⁴ for the statement that this speech-act has been performed; thus the happiness conditions of promising are the truth-conditions for the statement that someone has made a promise. Consequently the claim that the act in question has been performed can be defeated, for it may be rebutted by showing that one or more of the happiness conditions is not fulfilled. Unhappiness is of course not restricted to the performance of speech-acts. It is also characteristic of the performance of certain acts in other rule-determined practices which do not involve utterances, e.g., starting a race. Hence the claim that such an act has been performed can be defeated.

If the satisfaction of the happiness conditions could be established by merely "looking at the facts," then the application of rules would conform to the model of rule formalism. Defeasible concepts have at least two types of looseness which cast doubt on this simple model. This can best be clarified by constructing an imaginary constitution for promising. Consider a society in which one makes promises by placing one's right hand on one's heart and saying "I hereby promise. . . ." If the question arises whether an individual has or has not made a promise on a specified occasion, problems arise in two different dimensions. The first of these might be ascribed to the looseness of the boundaries of concepts in natural language. In our example, questions may arise as to what will count as placing one's right hand on one's heart. Must one take one's glove off? If one's hand is covered with tar, must one dirty one's shirt in order to promise? Or can one hold one's hand just in front of one's chest? The second set of problems arises because of the "openness" of the lists of happiness conditions. This can be manifested in different ways. First, one of the conditions may be unfulfilled, e.g., a marriage ceremony may take place

without a ring, or in our imaginary example, an armless man might be held to have promised without the stipulated gesture. Second, substitutes may be offered for one or more of the conditions; e.g., in our example a man with no right hand might attempt to promise by putting his left hand on his heart. Third, it is at least conceivable that unprecedented cases may arise in which in addition to the normal happiness conditions of the performance there is an unprecedented but relevant factor. In all three cases a decision is required as to whether a successful performance of the act in question has taken place, and in none of them is the issue one of determining the boundaries of concepts in natural language; for, e.g., placing one's left hand on one's heart cannot be construed as placing one's right hand on one's heart, but whether a promise has been made is still an open question. In all three cases the considerations which should determine a decision are the same, namely, the general purposes of the practice and the consequences of modifying the happiness conditions or their interpretation to include or exclude a new type of case. This type of looseness could be tightened up by introducing a closure rule to exclude unprecedented factors from consideration; in addition one could lay down that the specified conditions were jointly necessary and sufficient for the happiness of the performance. The fact that we do not tighten the rules and fix our interpretation of them in this way is not simply due to an ineradicable looseness of natural language, but to a policy decision.¹⁵ The general purpose is to avoid prejudging unusual cases and perpetrating injustices to particular individuals.

* * *

Having discussed some problems connected with the two premisses of the argument, it remains to examine the status of the conclusion. Both non-naturalists and their critics have taken for granted that the conclusion of a syllogistic normative argument of the type in question has a special status as "evaluative," "prescriptive," or a "value judgment,"¹⁶ which not only distinguishes it from statements bearing truth-values but also gives it a peculiar force. Our discussion in this section is

¹⁴ Unlike Austin we have excluded insincerity and abuse, i.e., what he calls "I-conditions," from what we call "happiness conditions." This seems clearly justifiable in such cases as promising. (See J. L. Austin, *How to Do Things with Words* [Oxford, 1962], esp. Lectures II and IV.)

¹⁵ See H. L. A. Hart, *The Concept of Law* (Oxford, 1961), ch. 7.

¹⁶ Hare, *The Language of Morals*, esp. pp. 92-93 and pp. 172-175; and Searle, *op. cit.*, pp. 43-44, 58.

directed against both parties to the controversy equally, for we should like to suggest that the apparent common ground between them rests on a pair of confusions. The usual argument obscures the two important distinctions, first, between acknowledging the existence of rules and accepting rules, and second, between the meaning of a sentence and the force of its utterance on a particular occasion.

Lewis Carroll in *Mind*, 1895, showed the necessity in any logical system for rules of inference in addition to the formation rules and axioms. Achilles could not compel the Tortoise to agree to the truth of the conclusion of a valid syllogism, though it agreed to the premisses. He could not see that to make an inference requires an inference rule in addition to *any* set of premisses. Vainly hoping that he could incorporate into the premisses the inference rule required, Achilles was led by the Tortoise into an infinite regress. An analogy to Carroll's famous argument may be constructed to demonstrate the distinction between acknowledging the existence of a rule and accepting it. Imagine Achilles and the Tortoise playing chess. Achilles checks the Tortoise's king; they both agree that the king is check. Achilles calls to mind the rule, "If the king is in check, it must be moved." He therefore announces the conclusion that the Tortoise's king must be moved. The Tortoise retorts, "Why on earth must I move my king? Will your logic grip me by the throat again? Or is it the meaning of the words this time (the logic of 'Check!')?" Does the Tortoise really have a problem? And must it be the same problem as in 1895?

On the previous occasion the Tortoise held Achilles in the grip of an infinite regress generated by the incorporation of the *modus ponens* rule of inference into the premisses. Here, however, the Tortoise accepts this rule of inference. Moreover, both players agree that the rule enunciated by Achilles is a rule of chess. Consequently both agree that it is a deduction by *modus ponens* from the rules of chess that the king must be moved. Now, a sentence such as "The king must be moved" can be used to make a true or false statement.¹⁷ (That one must drive on the left in England is a true proposition, one of the truth-conditions of which is the existence of a rule prescribing the relevant conduct.) But if the Tortoise agrees that this statement is true, must he not also agree that he is committed to moving the king? This is Achilles' puzzle. What the Tortoise is trying to teach Achilles is not the

lesson of 1895, but a new one. He aims to show Achilles the difference between acknowledging the existence of a rule and accepting it. One must distinguish the use of "ought" and "must" as modal operators in rules from their use to indicate or prescribe the best thing to do. Acknowledging the existence of a rule commits one to the truth of an inference by *modus ponens* that a particular case does indeed fall under the existing rule. There is an appropriate rule of chess; hence the Tortoise agrees that it is a valid deduction from the rules of chess that the king must be moved. Here the operator "must" merely indicates that there is only one move in this situation which is legitimate according to the rules of chess. But accepting a rule, unlike merely acknowledging its existence, commits one to viewing the fact that one is subject to the rule on a given occasion as a reason for conforming to it. That the committed agent ought to or must move the king follows from his acceptance of the rule that in these circumstances the king must be moved. But since the Tortoise need not accept the rule, he is not committed to moving the king. Rules will not grip a player by the throat, although his opponent may.

If this point is thought to be trivial, as indeed it is in the case of chess, its importance may be more dramatically manifest in numerous situations in life in which decisions must be made whether to follow or to break relevant rules. Without this distinction how could one describe the situation of the patriotic rebel under an iniquitous legal system? Surely he recognizes that according to the law he ought not to conspire against the regime, but he is not *thereby* committed to desisting from his activities. Neither word-usage nor the existence of rules commit individuals to courses of action. On the contrary, individuals, by accepting rules as their standards of conduct (and so constituting reasons for action) commit themselves to acting according to these standards.

The second point about the conclusion of the syllogism which is obscured by both parties to the controversy is the distinction between the meaning of the sentence and the force of its utterance. The recognition that the proposition "The king must be moved" is true in the Chess Derivation does not exhaust its philosophical significance. The uttering of a sentence which expresses such a (true or false) proposition may constitute the performance of diverse speech acts, e.g., enjoining, informing, exhorting, etc. It is a common assumption that the

¹⁷ See G. H. von Wright, *Norm and Action* (New York, 1963), pp. 194-196.

conclusion of the syllogism has a unique status as a "value judgment" or a "prescriptive" or "evaluative" statement. These characterizations appear to be references to different kinds of speech-acts. Prescribing seems to belong to the class of acts to which exhorting, enjoining, and commanding also belong, while evaluating is a member of a class containing acts such as grading and assessing.¹⁸ Consequently to call a type-sentence "prescriptive" is to assign it an illocutionary force. But illocutionary forces are properties of utterances. Hence, an illocutionary force could be assigned derivatively to a type-sentence only if it could be shown that any utterance of this type-sentence would necessarily have this particular force. It is doubtful whether any type-sentences satisfy this condition,¹⁹ and certainly sentences of the form "*X* ought to do *y*" do not fall into this category. They can be used to inform, criticize, give a verdict, agree, concede, illustrate a point, etc., and not only to prescribe or evaluate.

Thus if a refutation of non-naturalism must show that factual premisses can entail a prescriptive conclusion, it can never succeed. For entailment is a logical relation between propositions; therefore a proposition can "entail" an illocutionary force only in the sense that it entails a proposition which can only be expressed in sentences which on morphic grounds alone must have a unique illocutionary force whenever they are uttered.

On the other hand, naïve non-naturalism is

erroneous for the same reason in so far as it construes the truth about the structure of moral argument to be that an evaluative or prescriptive conclusion can be entailed only by premisses one of which is evaluative or prescriptive. For, if our argument is correct, no set of premisses can entail an evaluative or prescriptive conclusion.

* * *

This paper has been concerned only with the normative aspect of the naturalistic fallacy. We have tried to demonstrate the errors in one form of attempted refutations of non-naturalism. We first argued that the major premiss of a normative argument could not be construed as a definition or analytic truth even within practices, but is to be understood as a statement of a rule or principle. We then criticized the assumption that the second premiss was a "mere statement of fact." Finally we argued that neither the existence of rules, nor the truth of statements that particular cases fall under rules, generates commitments, and we questioned the "prescriptivity" of the deduced conclusion of normative syllogisms.

Although both non-naturalists and their critics seem to be seriously mistaken, the non-naturalists are nearer to the truth. For norms or standards of conduct are adopted by human choice, and commitment to principles or ways of life is not given either as part of the furniture of the world or as a consequence of lexicography.²⁰

The Queen's College

¹⁸ The first Austin termed "verdictives"; the second, "exercitives." (Austin, *op. cit.*, pp. 150 ff.)

¹⁹ J. L. Austin, *Philosophical Papers* (Oxford, 1961), p. 200: "in general of course . . . the use of any one sentence form does not tie us down to the performance of some *one* particular variety of speech-act."

²⁰ We owe thanks to Professor A. J. Ayer for his helpful comments on a draft of this paper.

VI. PERSONS, P-PREDICATES, AND ROBOTS

RAZIEL ABELSON

I. PREFACE

IN his essay "Persons," P. F. Strawson charted a promising new course through the dire straits of metapsychology, a course that keeps a safe distance between the Scylla of introspectionism and the Charybdis of physiological or behavioral reductionism. His main principle of navigation was his definition of psychological concepts as predicates that are both self-ascribable without observation and other-ascribable on the basis of "logically adequate behavioral criteria."¹

A. J. Ayer has recently argued that Strawson's "middle course" is illusory and leads to shipwreck on both sides.² He maintains that Strawson's notion of logically adequate behavioral criteria of *P*-predicates is incoherent, offering two main arguments to this effect, a dilemma and a *reductio ad absurdum*. The dilemma is this: Either the behavioral criteria of *P*-predicates are linked to them by definitions or they are not logically adequate. If the former, then reductionism is right. If the latter, then introspectionism is right. In either case, Strawson must be wrong.³

Ayer's second argument involves an imaginary experiment in which a child is brought up among robots and taught psychological language by a tape recorded voice that describes the robots as if they were persons, attributing the same psychological states to both robot and child when they make similar sounds and movements. The child then would learn to apply *P*-predicates to himself correctly and to the robots mistakenly, from which it follows that behavioral criteria of such predicates cannot be logically adequate, i.e., they cannot guarantee correct application.

I shall try to show that both of Ayer's arguments are unsound, the first because it oversimplifies the notions of behavior and logical adequacy, the

second because it conflates two problems that should be kept separate. I shall begin with the second argument, because the plausibility of the first depends upon the equivocation involved in the second.

II. THE CASE OF THE CLEVER ROBOTS

Ayer's description of his imaginary experiment in child rearing betrays a simplistic identification of human behavior with physical movements, that, if unchallenged, would justify the reductionism he wants to avoid. The child in his *Gedankenexperiment* is taught the use of psychological concepts by a recorded voice that attributes *P*-states to the robots when they make the appropriate gestures and sounds. But it is essential to Strawson's account of *P*-predicates that mere physical movements cannot be identified with human actions. Strawson says:

... we understand ... we interpret (bodily movements) only by seeing them as elements in just such plans or schemes of action as those of which we know the present course and future development without observation of the relevant movements. But this is to say that we see such movements as actions, that we interpret them in terms of intention. ... It is to say that we see others as self-ascribers, not on the basis of observation, of what we ascribe to them on this basis.⁴

Now if, to interpret physical movements as criteria of *P*-states of a person, one must "see" the initiator of these movements as a self-ascriber, then Ayer's robots cannot be mistaken by the child for real persons, because the tape-recorded voice cannot get him to see the robots as *self-ascribers*. Only the robots could accomplish that deception by simulating the behavior of self-ascribers, that is, by talking about themselves.

So the robots would have to teach the child

¹ P. F. Strawson, "Persons" in *Individuals* (Garden City, Anchor Books, 1963).

² A. J. Ayer, "The Concept of a Person," in *The Concept of a Person and Other Essays* (New York, St. Martin's Press, 1963).

³ *Ibid.*, p. 95: "But what exactly is meant here by saying that a criterion is logically adequate? Not that the evidence entails the conclusion, for in that case we should not stop short of physicalism. Not that the evidence provides sufficient empirical support for the conclusion, for then the reasoning is inductive; we are back with the argument from analogy. What is envisaged is something in between the two but what can this be? What other possibility remains?"

⁴ Strawson, *op. cit.*, p. 109.

- psychological language by ascribing *P*-predicates to themselves as they perform appropriate actions, e.g., saying "I am angry at you, I intend to spank you," and then spanking the child, just like real parents in the good old days. If the tape-recorded voice alone were to perform this pedagogical task, the child would soon come to regard the voice as the voice of a person (and correctly so), and the robots as merely the distant bodily organs of that same person, and he would be as close to the truth of the matter as the oddities of his eerie world permit. No doubt his idea of *another* person would differ from ours. He would think the world contains two persons: one (himself) with a head, two arms, a torso, and two legs connected vertically in that order, and the other a set of spatially separated metallic limbs controlled by a centrally located metallic head (the tape recorder). How easy or hard it would be for this poor child to develop filial feelings toward his mechanical "daddy" (or "mommy"?) is a matter of speculation. Recent experiments in which rhesus monkeys were nourished by robots indicated that simian babies can develop as strong attachments to dummies as to mummies. But we may doubt whether human children can be satisfied by milk without human kindness. In any case, the child's concept of a person would *not* be erroneous, for the recorded voice is the voice of a person. At worst, only his knowledge of biology would suffer.

But suppose Ayer were to amplify his imaginary experiment to meet this objection by postulating that a gifted engineer builds into the robots mechanisms that produce appropriate sounds on appropriate occasions so as to simulate rational and emotional discourse. Suppose then that the robots are programmed to say "I am angry" when the child strikes them and then to spank the child; to say "I love you" when the child kisses them and then to kiss the child and, in general, to make just the sounds that we ordinarily interpret as self-ascriptions of *P*-predicates, and to perform just the movements that we ordinarily interpret as motivated actions. Would the child then learn to ascribe *P*-states mistakenly to the robots while ascribing them correctly to himself?

This amplified hypothesis must be looked upon

with suspicion because it cannot even be stated without begging the crucial question as to whether genuine discourse and emotional behavior *can* be mechanically simulated to perfection, which is precisely the issue in contention between Ayer and Strawson. To begin by assuming that Strawson is wrong will not advance our understanding of the problem.

But whether Ayer's argument is question-begging is not the most important issue at stake. It is of more interest to decide whether his hypothesis, in the revised form, is a coherent one. Is it theoretically possible for machines to simulate human discourse and behavior with sufficient accuracy to provide a child with adequate and yet misleading paradigms of psychological states? Put in this way, the question smells of paradox. How could paradigms be both adequate *and* misleading? We are entitled to suspect that the hypothesis is, in fact, incoherent. Our problem will be to pinpoint where and how it goes wrong.

Now the issue as to whether robots can effectively simulate human behavior may be understood in two importantly different ways: (a) as the question how we know that a creature, *X*, is a person and (b) *assuming X* to be a person, how we know from his behavior that he is in a specific *P*-state.⁵ Ayer's preoccupation with the problem of other minds led him to interpret Strawson as if Strawson were offering an answer to the first question whereas, if I understand him, Strawson was only concerned with the second.⁶ In clarifying the first problem we shall be able to see what is wrong with Ayer's *Gedankenexperiment*, and in clarifying the second, we shall find a way out of his logical dilemma.

The first problem, how we know whether a creature is a person or a robot (or whether robots can perfectly simulate persons), may be seen in a new light by reversing Ayer's experiment of the child and the robots, and considering how we, who were presumably correctly trained in the use of psychological language, would decide if a creature who lands in a space ship from Mars is a person or just a robot. Let us set aside important, but for our purpose, distracting matters of morphology and biochemistry by assuming that our Martian looks enough like a human and enough unlike one so

⁵ I am suggesting here that being a person, i.e., being a subject to whom it makes sense to attribute *any* *P*-predicates, is a contextual presupposition for the ascription to an entity of a *particular* *P*-predicate. Cf. L. Wittgenstein, "Only of a living human being and what resembles a living human being can one say: it has sensations, it sees . . . is conscious . . ." (*Philosophical Investigations*, trans. G. E. M. Anscombe [New York, Macmillan, 1953], para. 281.)

⁶ Strawson explicitly warns: "Of course these remarks are not intended to suggest how the problem of other minds can be solved." *Op. cit.*, p. 109.

that its appearance provides no decisive clue, and that its substance is somewhat like but also somewhat unlike animal tissue so that we cannot be sure if it was manufactured or just grew. If we are to have criteria of its status as person or robot, they will have to be behavioral. Now suppose this creature moves and sounds very much like us. It does things that strikingly resemble walking, talking, responding to our speech and gestures, perhaps not exactly like anyone of us, but then neither are we exactly like each other. Would we automatically admit this creature into the Kantian society of rational beings? No doubt, some of us would; but others, more stubbornly sceptical, would suspect it to be a product of Martian engineering, able to simulate psychologically motivated behavior although cold and empty "inside," lacking the "internal" states of feeling and emotion that real persons have. Would there be any rational ground for denying such sceptics a logical right to be sceptical? I can think of no logical or semantical rule they might be accused of violating. In other words, we simply do not have established criteria for deciding whether *creatures other than human beings* are to be considered as persons subject to psychological description.

We are logically entitled to take any common properties of our human paradigms as criteria for applying the concept of a person to troublesome cases, but we are equally entitled to refuse the concept further application. A sensible and kindly man will treat a creature as a person if that creature exhibits many behavioral features in common with humans, so a sensible and kindly man will be considerate to dogs and Martian strangers. But Descartes was within his semantical rights in regarding animals as automata, and we would have the same right to suspect our Martian to be a robot. So Ayer is right that there are no generally established logically adequate behavioral criteria of being a person, but wrong in thinking that we *need* criteria. The problem of other minds is indeed insoluble (except by arbitrary decision) *when it is a problem*. But it is a problem only in very odd circumstances.⁷

Having separated the problem of criteria of being a person from that of criteria of specific *P*-predicates, and having granted Ayer the inconclusiveness of the former with respect to cases other than human beings, we can now return to the

lonely child among robot companions and see more clearly what is wrong with Ayer's description of the situation. For we can now distinguish two possible cases: Either (i) Ayer's robots are so clever that even we, who know the correct use of *P*-predicates, cannot be sure whether they are persons or robots, or (ii) they are clumsy enough to provide adequate grounds for identifying them as robots.

In the first case, the child would *not* be falsely or incorrectly ascribing *P*-predicates to the so-called "robots," for if they are so like us that we cannot distinguish them from our fellowmen, we have only Ayer's word for it that they are robots and, *ex hypothesi*, we have no reason to *take* his word for it. Indeed, to take his word for it is to beg the issue at stake by assuming that a person must have private mental states that can remain unknown to others. It is to assume that being a person is a matter of internally observable events. But since person-status is *presupposed* by the ascriptions of any *P*-predicates, the concept of a person must have a different linguistic function; it is not the name of a set of internal happenings, but rather a warrant for ascribing human characteristics (and especially *P*-predicates) to a subject so identified.

In the second case, where there are clear behavioral grounds for identifying the child's companions as robots, the child who is taught to describe their sounds and movements in words that *we* use to describe persons would not, as Ayer claims, be falsely ascribing psychological predicates to robots. He simply would not have acquired a vocabulary of genuinely *psychological* predicates. His word "person" would mean any creature that makes sounds and motions common to him and his robots, and his predicates, "angry," "sad," and the like would designate for him, not psychological states, but behavioral dispositions of which robots are as capable as people. Instead of having learned to apply psychological concepts to himself truly and to the robots falsely, as Ayer envisions the matter, the child would simply have learned a neutral language of the kind now popular among experimental psychologists and engineers.

Thus the success of Ayer's *Gedankenexperiment* rests on an equivocation between case (i) where robots are indistinguishable from persons and thus robots only by Ayer's arbitrary nomenclature, and case (ii) where the robots are plain robots. Once we note the incompatibility of the two cases, Ayer's

⁷ On the element of choice in deciding when to extend person-status to non-humans, cf. H. Putnam's illuminating discussions, "Minds and Machines" in *Dimensions of Mind*, ed. S. Hook, (New York, N.Y.U. Press, 1960), and "Robots: Machines or Artificially Created Life?," *The Journal of Philosophy*, vol. 61 (1964).

- counterexample to Strawson's thesis collapses, since it can be seen to depend upon the inconsistent hypothesis that both cases hold simultaneously.

III. THE DILEMMA AND THE WAY OUT

Having rid ourselves of the red herring problem of how we know that a creature is a person, we can turn to the more genuine Strawsonian problem of how, assuming X to be a person, we can know by observation and with logical certainty what psychological state X is in. Strawson claims that it is of the essence of a P -predicate that it can be ascribed to others on the basis of logically adequate behavioral criteria. We can now take up Ayer's first argument against Strawson, namely, the dilemma that if behavioral criteria entail P -states then P -states are nothing but behavioral tendencies and reductionism is right, while if the criteria merely provide inductive evidence for P -states, then they are not logically adequate and introspectionism is right. According to Ayer, there is no third possibility in sight. In what follows I shall attempt to sketch out just such a third possibility.

Ayer seems to think of the relation between a state s and its criterion C as either the law-like but contingent relation between two regularly associated facts (e.g., fever and bacterial infection), or as the analytic relation between a concept and its defining properties (e.g., being a bachelor and being unmarried). Both models are too simplistic to do justice to the linguistic role of a criterion.

The first step in finding a more adequate explanation of the relation between criterion and state is to distinguish between the unrestricted logical implication that holds between a term and any part of its definition (as in bachelor-unmarried) and the contextually limited implication that holds only under standard conditions between a dispositional concept and its operational criteria. This relation was made formally precise by Rudolf Carnap in "Testability and Meaning."⁸ Carnap suggested that dispositional predicates like "soluble in water" require a special and incomplete mode of definition which he called, "reduction sentences." A reduction sentence contains a major conditional clause that specifies the standard test conditions (e.g., the condition that X be placed in water) for the implication asserted in the conse-

quent clause between the disposition (soluble) and its criterion (dissolves). Formally,

$$(x)[Q_1(x) \supset (Q_2(x) \supset Q_3(x))]$$

where " Q_1 " designates the test condition of being placed in water, " Q_2 " designates the criterion, dissolves, and " Q_3 " the dispositional property of being soluble in water. Gilbert Ryle, in accounting for psychological states as behavioral dispositions, seems to have been guided by a similar model.⁹ But the model is still too simple. For the open texture of psychological predicates like "angry" prevents us from listing any observable conditions under which " x is angry" entails or is entailed by one or more observable responses such as " x shouts" or " x attempts to strike y ," etc. To specify such standard conditions we would need invariant laws of psychology and it is no secret that we have no such laws.

But Carnap's notion of dispositional implication at least sets us on the right track in exhibiting contextual limitations on the scope of the implication. Such limitations are bound to be more complex for P -predicates than for physical properties, for two reasons: (a) the lack of invariant laws and (b) a factor which, I believe, explains why there cannot be invariant laws of psychology, namely, the semi-evaluative function of P -predicates.

Take, as a test case, the predicate, "angry at y ." We have already noted the folly of assuming that there is any single response such as x striking y , or any finite disjunctive set of responses (such as x striking or shouting at or insulting or shooting or . . . y), entailing or entailed by " x is angry at y ." It is conceivable that x might be angry at y and do none of these things and that x might do any of these things without being angry. The type of response that we may reasonably interpret as anger depends on our evaluation of the circumstances. Suppose that x is the clandestine mistress of a public official, y , and that y has cast her off. On encountering y in a public place, x embraces and kisses him in order to embarrass him. Seen in this light, kissing qualifies as a normal anger response because of the special circumstances. On the other hand, suppose x is an employee of y , y fires x , and x kisses him. The kiss in this case is surely not a normal expression of anger. We would have to inquire further to find out just what psychic state it manifests. So the relation between P -predicates and their criteria is limited by two general condi-

⁸ R. Carnap, "Testability and Meaning," *Philosophy of Science*, vol. 3 (1936), Part I.

⁹ Cf. G. Ryle, *The Concept of Mind* (New York, Barnes & Noble, 1960), ch. V.

tions: (1) that the eliciting circumstances be of a kind appropriate to the *P*-state in question and (2) that the state be manifested by normal responses, i.e., responses appropriate to the eliciting circumstances. In the case of anger, we indicate the relevant circumstances by saying that the agent was provoked, and we signify the appropriateness of the response by saying that he responded in a normal way, or was in a normal state of mind.

Applying these considerations to the problem of formulating logically sufficient criteria for *P*-predicates like anger, we need a formula like a Carnapian reduction sentence, but more suitable to the open texture and evaluative force of psychological concepts. Anthony Kenny's illuminating analysis of the criteria of emotions helps make clear what must go into such a formula. Kenny identifies three kinds of criteria of emotions, all of which apply in simple, paradigm cases, but any of which may be absent in special cases: (a) provocation, (b) physiological symptoms, and (c) behavioral responses.¹⁰ Kenny points out that while none of these is a necessary condition for the correct ascription of an emotion, their conjunction is a sufficient condition. Following this lead, we might construct a reduction formula or meaning postulate for "*x* is angry at *y*," by specifying the general restriction of *P*-predicates to persons, and the three suggested behavioral criteria, as follows:

If *x* and *y* are persons and *y* provokes (does something bad to) *x*, then, if *x* is agitated (flushes, pales, clenches his teeth, perspires, etc.) and *x* does or tries to do something appropriately bad to *y* (retaliates), then *x* is (necessarily) angry at *y*. Formally,

Necessarily: $(x)(y) \{ (Px \& Py \& Pryx) \supset [(Sx \& Rxy) \supset Axy] \}$

In criticism of this formula it might be argued that, while the formula gives conditions that are *empirically* sufficient for (i.e., have a high correlation with) the state of anger, it does not provide *logically* sufficient conditions, because we can conceive of cases where all the antecedents of the formula hold and the consequent does not. Such counter-instances, it will be conceded, are fairly rare, for if they were frequent, it is hard to see how anyone could learn to identify psychological states at all. Nonetheless, the argument will go, no finite list of conditions can logically rule out all possible exceptions, and therefore no finite list can provide logically adequate criteria.

Now it seems to me that, plausible as this objection may sound, it *must* be wrong, because we not only learn to apply *P*-predicates to simple cases where all the above criteria work, but we also learn how to identify the exceptional cases where our criteria break down. This second step makes all the difference between a naïve and a sophisticated understanding of psychology. Now if we can learn how and when to spot exceptions, then the exceptions themselves must come under general rules, and we can build these rules into our formula. Let us try to classify the *types* of possible exceptions, and perhaps we will find either that our formula has already excluded them, or that a slight modification of the formula will serve that purpose. For the ascription of emotions such as anger, the exceptions to our criteria can, I think, be subsumed under three general types, as follows: (1) cases where alternative emotions may explain the observed behavior; (2) cases where purely rational considerations, i.e., practical or moral reasons rather than emotional states, may explain the observed behavior, and (3) cases where the emotion is simulated rather than genuine. Let us then consider whether our formula has already ruled out these cases and, if not, whether it can be adequately modified.

Type 1 (alternative emotions): *x* may show agitation and respond aggressively, yet be motivated by some state other than anger, for example, fear. In such a case, the factor that distinguishes fear from anger (or from a mixture of both) may be found in subsequent responses of the agent. If *x* strikes *y* in fear rather than in anger, he will desist when he finds that *y* is stronger than he or when the danger *y* presents to him is removed. But if he is acting in anger, he will continue to react beyond the needs of self-defense and even contrary to such needs.

Type 2 (non-emotional reasons): *x* may respond aggressively to *y* for reasons that are, in themselves, non-emotional in either of two ways—*x* may be acting to achieve some practical goal such as the deterrence of further aggression by *y*, or for the sake of some moral principle such as professional duty or social justice. But these cases are already ruled out in our formula by the condition of symptomatic agitation. It may be argued that all responses to provocation, no matter how rational their justification, are attended by some degree of agitation. But if so, then all such responses are, to some degree, emotional, and there can be no counter-instances of this type to our formula,

¹⁰ A. Kenny, *Action, Emotion and Will* (New York, Humanities Press, 1963), pp. 67 ff.

- which purports to give *sufficient*, but not necessary conditions for ascribing anger. There is no need to exclude the possibility that an angry person has a good practical or moral reason for retaliating against those who provoke him.

Type 3 (simulated emotion): What of those cases where x is merely simulating anger, like the politician whose constituents expect a proper fury at high taxes, communist victories, or crime in the streets, or the Method actor who convinces himself in order to convince his audience? This kind of exception literally proves our rule, for how could it be possible to stimulate a state if there were no adequate criteria to take advantage of? What could one *do* to simulate anger, if the state of anger were logically independent of behavior? The very possibility of pretense entails, not the logical insufficiency of criteria of the state that is simulated, but rather the possibility of our being mistaken as to whether the criteria are in fact satisfied. A counterfeit bill is not one that satisfies the criteria of a bank note yet somehow fails to be one, but one that superficially appears to satisfy, yet on closer inspection can be seen not to satisfy the appropriate criteria. Similarly, if x is simulating anger, then he is not really agitated or not really trying to injure y , and sufficiently close inspection should reveal that one or both of these criteria are not in fact satisfied.

I can imagine no other types of possible exceptions, real or apparent. This, of course, is not to say that no other exceptions can be found. Still less can I claim to have proven that *all* P -predicates have logically adequate behavioral criteria, even if my formula were conceded to work for the particular predicate, "angry at." But I doubt if Strawson meant to claim that *all* P -predicates have adequate criteria. Predicates signifying long-term states or traits, e.g., "loves," "is honest," "is ambitious," probably do not have adequate criteria for the reason that no time limits can be agreed upon for their correct ascription. A man may act like a lover, may behave impeccably, or may pursue public honors for a time and then stop. Are we to say that he loved but no longer loves, that he was honest for a while but became corrupt, that he had, but then lost, ambition or should we say that he was not really in love, not

really honest, not really ambitious? There simply are no precise rules governing such a decision. But as I understand Strawson, he claims only that those P -predicates that can be ascribed to oneself without observation (and thus with absolute authority) must also be ascribable to others on the basis of logically adequate behavioral criteria. Now long-term attitudes or traits can no more be ascribed to oneself with non-observational authority than they can be ascribed to others with empirical certainty.

At the other end of the psychological spectrum, predicates that signify very transient states, e.g., feelings, sensations, and moods, are particularly easy to ascribe by means of behavioral criteria. A groan subsequent to an injury, a radiant smile after winning a contest are (setting aside the special problem of simulation with which we have already dealt) logically sufficient criteria of pain and pleasure. My reason for choosing the predicate, "angry at," as a test of Strawson's thesis was that it falls just about halfway between the easy short-term cases and the impossible long-term cases. I think that Strawson would have to grant that there are no logically adequate behavioral criteria for very general and long-term P -predicates, but the concession would not endanger his thesis. Persistent attitudes like love and character traits like honesty can be explicated, *a la* Ryle, as tendencies to experience certain short-term states like affection or desire, to perform certain activities like doing favors or telling the truth, and to manifest agitation on appropriate occasions—in brief, they may be explicated as tendencies to definite P -states, rather than as definite P -states in their own right.

It has not been my intention to attempt the impossible task of proving that *all* P -predicates have logically adequate behavioral criteria, nor even of proving this for the single predicate, anger. I have tried only to sketch out the schematic form in which adequate criteria could be articulated, and to fill out the form plausibly for the predicate, anger, in such a way as to show a possible escape from Ayer's dilemma. I do not pretend to have proven that Strawson is right, but only that it is possible that he is right. Nor am I at all confident that Strawson would recognize his baby in all these new clothes.

VII. PLEASURE AS AN END OF ACTION

JAMES D. WALLACE*

MY aim in this paper is to describe some of the philosophically important ways in which things done for pleasure differ from things done for certain other reasons. The main contention of this paper is that things done for pleasure are paradigms of "free" acts, but that the same features of things done for pleasure which make these paradigms of "free" acts also make the agent particularly vulnerable to certain sorts of criticism, should the agent in doing something for pleasure do something wrong. This is so because the very fact that someone does something for pleasure puts a certain range of excuses and justifications beyond his reach, should he be called upon to defend his actions.

I

Although people sometimes enjoy doing the things that they do in order to fulfill their needs and obligations, generally it is when their actions are *not* prompted by such considerations that they are said to act for pleasure. It is characteristic of things done for pleasure that they are "spare-time" activities—activities which occupy the interstices left by the duties and necessities of life. This is connected with the point that the things that people do for pleasure are not apt to be matters of urgency or importance. In fact, there is a definite sense in which things done for pleasure are *less* important than things done to fulfill needs and obligations. What a person does for pleasure may be important *to him* in the sense that he is willing to go to considerable trouble to do this sort of thing. However, there is a kind of importance which cannot be claimed for things done for pleasure, and this is shown by the following considerations.

If a person were called upon to *justify* his neglect of his own needs or his failure to fulfill a certain obligation, he could not do this simply by establishing that he did something for pleasure instead of attending to these other matters. This would be no justification at all. It might be that this person

actually regards his pleasure as more important than the neglected need or obligation, but this will not justify his neglecting these things—that is, it will not show that what he did is perfectly all right. One *might* in some cases be inclined to look less severely upon someone's neglecting his needs or obligations if one knew that he had neglected these matters to engage in some particularly rare pleasure. However, this would not justify his neglect in the sense of showing that there was nothing wrong with what he did. Moreover, it may be that this is an excuse only because in pleading the rareness of the pleasure, one implies that this is a *rare lapse*. We sometimes look less severely on momentary weaknesses and occasional lapses on the ground that such lapses are rare. If someone's lot in life were a particularly burdensome one—if circumstances were such that all of his time was taken up with onerous tasks which his needs and obligations required of him, these circumstances might count as mitigating circumstances should this person neglect his needs or duties to do something for pleasure. However, here again it is not just that the person does something for pleasure that constitutes his defense. Rather, it is the oppressive character of his lot in life. It seems *unfair* that anyone should be so burdened, so we might make special allowances in his case. At any rate, establishing that one neglected to do something, which for moral or prudential reasons one should have done, to do something else for pleasure does not justify or even tend to excuse the omission. There might be other circumstances in such a case that one might cite as mitigating circumstances, but citing the fact that one did something else for pleasure—even pointing out that the pleasure is more important to one than the neglected need or obligation—is no defense at all.¹

One could justify failing to fulfill an obligation or neglecting a need if one could establish that what one did instead was actually more important

* I wish to thank Professors Charles E. Caton and B. J. Diggs for their helpful comments on a draft of this paper.

¹ On the distinction between excuses and justifications, see J. L. Austin's "A Plea for Excuses" in *Philosophical Papers*, ed. J. O. Urmson and G. J. Warnock (Oxford, 1961), pp. 123-152; especially pp. 123-125.

no means of identifying which piece is the king. Similarly, if merely told that the king is the piece which can move only one square at a time in any direction, we must still find out which piece it is that can move this way. Rules are pointless unless it is possible to distinguish the individuals to which they apply from those to which they do not. Consequently to offer a rule as a definition of the piece to which it applies neither individuates the piece nor provides a criterion for the application of the rule. As Wittgenstein pointed out, "if one were to say 'The king in chess is *the* piece that one can check', . . . this can mean no more than that in our game of chess we only check the king." It would be a mistake to suppose that we had a concept of "check" which we could use to determine what is and what is not a king.

Of course, in one sense, the rules which apply to the king serve to differentiate the king from every other chess-piece. The rules about the king differ from those about the queen, and consequently the answer to the question of what distinguishes the king from the queen *may* be given in terms of the rules that apply to each. But in neither case do the rules specify to which pieces they apply. And in another sense of the same question, an answer cannot be given in terms of the rules. For, if we have explained to someone what the rules of the game are, and he wants to begin to play, his question would be a demand for a definition of the king in terms which he knows how to interpret, i. e., he wants an interpretation of the rules, and being subject to rules is not a candidate for such an interpretation. It is only by confusing these two different senses of the question "What is a king?" that one jumps to the conclusion that it is one part of the meaning of the word "king" that the king is subject to such and such rules and that it is another part of the meaning of the word "king" that the king is the piece that stands on *E1* when the pieces are in their initial positions. But these two "parts" of the meaning are very different and only confusion can result from combining them under the one heading "the meaning of 'king'."

If we wish to individuate the piece to which the rules about the king apply, various possibilities are open. The most convenient method of giving an interpretation would be to specify what is to count as a piece (e.g., a counter) and to differentiate the pieces by reference to where they stand in the initial

position and by their "color." In this constitution the initial position is defined as the arrangement of the sixteen pieces of each "color" in two rows at the appropriate ends of the board. Now the white king, e.g., may be defined as the piece which stands on *E1* in the initial position. This constitution enables chess to be played with physically diverse sets or even with undifferentiated draughts without making any difference to the game (apart from confusing uninitiated spectators).

But even if the king cannot be exhaustively defined in terms of any *one* of the rules that apply to it, perhaps it can be *partly* so defined. Is it possible to construct a definition in which one or more rules will appear as components? A suggestion would be that the king be defined as the piece in chess which satisfies the following two conditions:

- (1) It stand on *E1* in the initial position.
- (2) If it is in prise, it must be moved.

If one of these conditions is not to be discarded as redundant, it must be conceivable that a piece in chess could satisfy one but not both of these conditions; in short, they must be *independent*. Yet it is inconceivable that there be a piece in chess which could satisfy either condition without satisfying the other.⁸ If there were a piece which when in prise had to be moved, but was not the king, a conflict of rules might occur which would cause the game to collapse (or multiply the possibilities for stalemate or checkmate). For both this piece and the king could be put simultaneously in prise, and it is impossible to make two moves at one turn. Similarly two different pieces cannot stand on the square *E1* in the initial position. Finally, it is not possible that there be no piece that satisfies both conditions, since in the absence of the king the game would have no point and consequently no conclusion. The obvious reason for the dependence of these conditions is that they are both parts of a constitution; the second condition is a rule of the system; the first is one of the formal definitions of the system, and, given the interpretation of the primitives, gives the criterion for identification of the piece to which the rule applies.

Moreover the suggested definition gives an answer to the two quite different questions, ambiguously expressed by the question "What is the king?" or "What is the meaning of 'king'?", viz. "What are the rules governing the movements of the king?" and "To what piece do these rules

⁷ L. Wittgenstein, *Philosophical Investigations* (Oxford, 1953), § 196.

⁸ This would be false if we were defining, e.g., a Queen's Rook or a King's Bishop. If, however, we are to define a Rook of a Bishop, the first condition would consist of a disjunction of two positions, and the argument would remain true.

apply?" But this comprehensiveness, which might appear a virtue in a definition, is a vice. Compressing the answers to these two questions into one definition is misleading in that it suggests that there is one question to the answer of which the two components of the putative definition are each necessary and only jointly sufficient conditions. This in turn suggests that these two conditions are independent, which we have seen not to be the case. The statements "If it is in prise it must be moved" gives a complete answer to a request for one of the rules which uniquely determines the king, but provides no answer to the question whether this piece is the king. This question is exhaustively answered by the statement "The king is the piece that stands on *E1* in the initial position" and stating the rules which apply to the king does nothing toward answering this question. Only if the compressed "definition" is an answer to a request for an explanation of the constitution of chess as a board game is one or other of the components not redundant or irrelevant. For the one type of component specifies the rules of chess, and the other furnishes us with a definition; the conjunction of the two together with an interpretation for the primitives exhaustively defines chess as a board game.

While a constitution may consist of both rules and definitions, as in chess, it is at least sometimes possible to write the definitions into the rules. This is accomplished simply by eliminating the defined terms by substituting the *definienda* for their occurrences. The function of definitions, when they occur in a constitution, is to pick out the individuals to which the rules apply and the occasions on which they do so. If these conditions of applicability are, so to speak, "written on the face" of the rules, then there is no need for such definitions. This situation is realized when it is considered that the problem of their interpretation is regarded as trivial. One example of this might be setting out the permitted moves of chess pieces beside pictures of the pieces governed by these rules.⁹ Another slightly different example might be the following schematic constitution for promising. Let "to pro-

mise to ϕ " be defined as "to utter in certain circumstances the words 'I promise to ϕ '." And let the rule of the constitution be that anyone who promises to ϕ ought, *cet. par.*, to ϕ . Then this rule and this definition could be amalgamated into a single rule to form a different, but equivalent, constitution for promising. It would consist only of the rule that anyone who utters in certain circumstances the words "I promise to ϕ " ought, *cet. par.*, to ϕ . (Here it is assumed that the utterance of "I promise to ϕ " is identifiable.) The conditions of its applicability are now "written on its face" in that further definitions would be otiose.¹⁰

This discussion of the relation of rules to definitions in constitutions reveals one of the errors in the attempted refutation of the naturalistic fallacy, viz., the interpretation of the major premiss in a normative argument, i.e., a rule or principle, as a definition or analytic truth.¹¹

* * *

Another possible source of error is the oversimplification involved in saying that the minor premiss of the argument is "straightforwardly empirical,"¹² or a "mere statement of fact." The minor premiss is intended to specify a particular case as falling under the rule (i.e., the major premiss). The problems here raised are related to the application of rules. If the doctrine known as "rule formalism" in legal philosophy were tenable, then the only difficulties involved in applying rules would be due to the looseness of boundaries of concepts in natural language.¹³ The situation is, however, much more complex than this. Difficulties about the application of rules to actual cases arise in respect to each of the elements of the rules, e.g., whether the particular person belongs to the class of persons specified as subject to the rule or whether a particular act performed is of the class of acts specified as permitted, required, or forbidden by the rule. In the following discussion we shall only briefly consider the performance of certain types of acts under rules, though an analysis of the

⁹ L. Wittgenstein, *The Blue and Brown Books* (Oxford, 1958), pp. 13-14.

¹⁰ L. Wittgenstein, *Philosophical Investigations*, p. 217.

¹¹ Failure to see this point lies at the root of the inadequacy of Hare's criticism of this form of argument (*Language of Morals*, p. 44).

¹² Searle makes this claim for one particular argument, but thinks that it can be generalized: "however loose the boundaries (of the concept of a promise) may be, and however difficult it may be to decide marginal cases, the conditions under which a man who utters 'I hereby promise' can be correctly said to have made a promise are straightforwardly empirical conditions." (Searle, *op. cit.*, p. 45.) The statement that these conditions obtain, and its consequence that a promise has been made, are straightforwardly empirical.

¹³ Searle would seem to adhere at least *de facto* to this doctrine.

• than the neglected need or obligation, and this might be done by showing that one was trying to fulfill a more important need or a more important obligation. However, if one did x for pleasure, and one had no more reason than this for doing x , then one cannot have been doing x to fulfill a need or obligation. No one needs to do x or is obligated to do x simply because he gets pleasure from doing x . That is, one could not sustain the claim that one needs to do x or that one is obligated to do x by the fact *alone* that one enjoys doing x , any more than one could sustain such a claim by the fact alone that Betelgeuse is in the constellation Orion. Something more must be added to the fact that S gets pleasure from doing x if we are to be entitled to say that S needs to do x or that S is obligated to do x . There might be cases in which someone could establish that he needs to do x or is obligated to do x *because* he enjoys doing x , but these will not be cases where the agent does x just for pleasure. For example, one might establish that one's health is in imminent danger of deteriorating unless one does something that one enjoys. This would support the claim that one needs to do x because one enjoys doing x . However, in such a case, the agent would be seeking the pleasure of doing x for the sake of the effect he thinks this will have on his health. Hence, this is not a case of doing x just for pleasure, since the agent can truly be said to be doing x for his health. Moreover, there is not yet any ground for adding to the explanation that the agent, also doing x for his health, is, also doing x for pleasure. That someone does x for pleasure implies that he seeks the pleasure of doing x for its own sake (as opposed to seeking the pleasure of doing x because of some further effect he expects this to have).² So far, we do not know that the agent who is doing x which he enjoys in order to preserve his health is also seeking the pleasure of doing x for its own sake, and, consequently, there is no ground for saying that in addition to doing x for his health, he is also doing x for pleasure.

One cannot be doing x to fulfill a need or obligation if one is simply doing x for pleasure and for no other reason. In order to sustain the claim that one is doing x to fulfill a need or obligation, one must establish that there is some reason for doing x beyond the fact that one enjoys it, and one must establish that one is doing x *for this reason*. Having established this, one will have established that one

is not doing x just for the pleasure of it. Consequently, if one did x just for pleasure, one cannot honestly claim that one did x on that occasion to fulfill a need or obligation.

I cannot justify my being remiss in some obligation or my neglect of my own needs simply by claiming that I did something else for pleasure instead of attending to these matters, and this would be so even if I regard my pleasures as more important than the neglected needs or duties. The sort of importance which is required for such a justification cannot be claimed for something done for pleasure *qua* act performed for pleasure. A person who regards his pleasures as more important than his needs and/or obligations to the extent that he frequently neglects the latter for the former is by this very fact defective in character. There is no one term for this sort of defect. Depending upon what sorts of duties and/or needs the person neglects, the sorts of pleasures he pursues, his own attitude toward this, and other factors, he can be described by such terms as: *dissipated, frivolous, self-indulgent, weak, lax, soft, imprudent, intemperate, selfish, untrustworthy*. The terms in this list refer to character defects—traits of character which are (more or less) bad to have. Even if an individual rarely neglects his needs or duties for his pleasures, doing so even once—except, perhaps, in very special circumstances—is properly characterized as acting wrongly or acting unwisely, although such isolated lapses are not necessarily enough to mark the agent as defective in character. One cannot justify neglecting a need or obligation by claiming that one did something for pleasure instead, because neglecting needs or obligations to do something for pleasure is itself a sort of wrong action. It follows from this that one's needs and obligations ought to take precedence over the things one does for pleasure, and in this respect the things that one does to fulfill one's needs and obligations are more important than one's pleasures.

Things done for pleasure alone and those done to fulfill needs and obligations constitute two distinct categories of action. Insofar as an act is performed for pleasure it cannot be done to fulfill a need or obligation. Moreover, the whole atmosphere which surrounds things done for pleasure is apt to be quite different from acts in the other category. Things done for pleasure—insofar as they are done for pleasure—do not have the air of

² Not all motive explanations of the form " S did x for m " imply that S sought m for its own sake. For instance, someone might raise mushrooms in his spare time for profit, not because he is interested in getting or having money for its own sake, but because he wants extra money to travel. The logic of "for pleasure" differs in a number of respects from that of "for profit."

urgency or imperativeness that often accompanies acts aimed at fulfilling needs or obligations. The latter acts have a greater importance, a greater moral and prudential importance than things done for pleasure. This is not to say that there is anything wrong with acting for pleasure *per se*. Rather, this means simply that needs and obligations have a sort of precedence over things done for pleasure.

II

Although in general it is wrong to neglect one's needs and obligations to do things for pleasure, there is no impossibility in someone's knowing this and still preferring to pursue his pleasures and neglect these other matters. Not only is such behavior possible, but it would also be paradigmatic of acting freely. Someone who is doing x for pleasure alone cannot be said to be *forced* by duty or necessity to do x . Moreover, someone who on a particular occasion did x for pleasure cannot upon that occasion have been *coerced* into doing x . A person, P , is coerced into doing x only if some person (or persons), C , leads P to think that C will cause something bad to happen to P , if P does not do x , so that P does x in order that this bad thing will not be done to him. We might get P to do x by convincing him that he would enjoy doing x , and we might thereby be *enticing* P into doing x ; but we would not be coercing him. For one thing, we would not be making P think that we would make something *bad* happen to him if he did not do x .

There is still another way in which one cannot be forced to do x , when one does x for pleasure. Extremes of appetite such as extreme hunger, thirst, or sexual desire, such strong emotions as fear, and compulsive cravings such as the cravings of an addict are sometimes said to force or drive people to act. There is no force of this kind when one acts for pleasure. Aristotle remarked that "self-indulgence is more like a voluntary state than cowardice" because self-indulgence is "actuated by pleasure" and cowardice by "pain."³ This remark is explained no further, but it does suggest that acts which are actuated by pleasure are "more voluntary" than those actuated by "pain," and there is a sense in which things done for pleasure are "more voluntary" than *certain* acts brought about by "pain." The particular form of "pain" which actuates cowardly behavior is fear, and fear can

force one to act contrary to what one most wants to do, and fear can force one to do something which one knows is in no way worth the evil it involves. There is nothing like this force when one acts for pleasure. One might do something for pleasure, knowing all along that one is being foolish, selfish, self-indulgent, etc.—even knowing that one will be sorry later. However, nothing could be said to be *forcing* one to do this.

One who acts for pleasure cannot be driven or forced to act in the way in which one might be driven or forced to act by fear, and it is easiest to see why this is so in the following sort of case. A person who is especially fond of the taste of lobster might on some occasion decide to eat lobster, knowing all the time that it will not agree with him and that he will be sorry later that he ate the lobster. Assume in this case that there are other wholesome foods readily available, and that his decision to eat lobster is based simply upon the fact that he enjoys eating lobster. It could not be said that anything forces him to eat lobster in such a case, because, given the fact that he eats lobster simply because he enjoys eating it, it follows that he wants to eat the lobster *more* than he wants to avoid the discomfort he expects will result. If a person eats lobster, knowing that as a direct result he will be uncomfortable sometime later, and if he decided to eat lobster simply for the pleasure of eating it, then this implies that he thinks that the pleasure of eating lobster is *worth* the discomfort he expects will follow—or at least it implies that he thinks that the chances of this being so are good enough to make it *worth* risking the discomfort. His deciding to eat a food which he knows will not agree with him, simply because he enjoys the taste of that food, shows *how much* he enjoys it (or how much he expects to enjoy it). There is nothing here that forces him to do what he would rather not do. No matter how terrible he expects the consequences to be, if he eats the lobster simply because he likes the taste, it will follow that he thinks that the pleasure of eating lobster is worth the pain that he anticipates, or at least that he thinks that the chances of this being so are good enough to make it worth the gamble.

On the other hand, fear can force one to do something that one knows is not worth the pain or other evil inherent in it, and in this way fear can force one to act contrary to what one most wants to do. For instance, a person might know that

³ *Nicomachean Ethics*, 1119a22–25. The translation is that of W. D. Ross in *The Basic Works of Aristotle*, ed. R. McKeon (New York, 1941).

minor surgery would correct a painful condition and prolong his life, and yet fear might force him to shrink from the operation itself. He might be certain that the benefits of the operation are well worth the immediate pain and risk it would occasion, and yet his fear might prevent him from submitting to the operation. The fact that he does not submit to the operation because he is afraid does not imply that he thinks that the benefits of the operation are not worth the discomfort and risk of the operation. Nor does it imply that he would rather forego the discomfort and danger of the operation and suffer the consequences than risk the operation and be cured.

When someone does x for pleasure, expecting unpleasantness as a result, the anticipated unpleasantness can be characterized as the trouble he is *willing* to face in order to do x . The amount of unpleasantness which he is willing to face to do x for pleasure indicates *how much* he enjoys or expects to enjoy doing x —i.e., what it is *worth* to him. He is not forced to take what he regards as a greater evil for a lesser good, because his doing x for pleasure shows that he thinks that the pleasure will be worth the unpleasantness he expects, or that the chances of this being so are good enough to make it worth the risk. It may turn out *later* that the agent discovers that what he did is not worth the ensuing unpleasantness. This, however, indicates that he was *mistaken* in his original estimate, and not that he acted contrary to his estimate of the relative worth of doing x or abstaining.

This, I think, is what lies behind Plato's contention that "being overcome by pleasure" is always the result of "ignorance."⁴ One is "overcome by pleasure," according to Plato, only if one does something, x , for pleasure, and only if one thereby takes "the greater evil for the lesser good"—that is, only if the pleasure of doing x is "unworthy" of the evil it involves.⁵ However, we have seen that one logically cannot do x for pleasure, believing all along that the pleasure of doing x will be in no way worth the pain or other evil it will involve. Consequently, if someone does x for pleasure, and if he discovers afterward that the pleasure of doing x is in no way worth the evil involved, we can conclude that he did not know this at the time he undertook to do x for pleasure. This is so because that one did x for pleasure implies that one did x thinking that it would be worth whatever consequences one anticipated.

It cannot be argued in this way that being overcome by fear is always the result of ignorance. Someone's fear might prevent him from doing what he most wants to do or it might force him to do something which he knows is in no way worth the evil it involves. This is what takes place in at least some instances of cowardly acts. Courage is not simply the ability to decide what is best or most desirable in the face of something fearful. Courage involves the ability to *act* upon such decisions.

It might be objected that at best I have shown only that when one does x for pleasure, one cannot be forced to do x *contrary* to one's estimate of the relative worth of doing x and abstaining from doing x . It does *not* follow from this, the objection might continue, that someone cannot be forced to do x when he does x for pleasure—all that follows is that he cannot be forced *contrary to a certain sort of estimate*. It might be suggested that in a case where someone does x for pleasure, his desire to do x might force him to do x , if the desire were strong enough. People are sometimes driven or forced to act by strong desires. All the same, it is important to notice that different things can be included under the heading of "being driven by strong desire."

Individuals are sometimes said to be driven by a strong desire to get ahead, to become wealthy, or to gain power, where this means simply that they are *willing* to go to extraordinary lengths to gain these ends. Despite the appropriateness of the term "drive" in such cases, it is not appropriate to regard this kind of strong desire as something which *forces* or *coerces* the agent. That is, the agent in such cases is not being made by his strong desire to act against his will, or unwillingly, or involuntarily. However, there is another sort of strong desire which can force or coerce. The strong desires connected with appetites and other felt cravings—for example, extreme hunger or the craving of an habitual smoker deprived of tobacco—sometimes force people to act. In these cases, however, strong desires can be said to *force* a person to act only if it is *not* the agent's view that the action prompted by his strong desire is worth any bad consequences he foresees. If someone believes that the pleasure of smoking is worth the risk to his health, and there are no other reasons why he should not smoke at the moment, then, although the craving for tobacco may prompt him to light up, he is not being forced. Such desires as these *force* or *coerce* one to act only

⁴ *Protagoras*, 357c-357e. The translation is that of B. Jowett in *Plato's Protagoras*, ed. G. Vlastos (New York, 1956).

⁵ *Ibid.*, 355d-355e.

if the action prompted by the desire does not coincide with the way in which one's will is inclined. The belief that smoking is worth its drawbacks puts one's lighting up in the category of actions which are to be contrasted with those which result from force or coercion.

It is true that if *S* does *x* for pleasure, then *S* must have had a desire to do *x* (i.e., he must have *wanted* to do *x*). However, *S*'s desire to do *x* cannot force or coerce him to do *x*, if he does *x* for pleasure. This is so because his doing *x* for pleasure *guarantees* that his will is so inclined—that is, it guarantees that *S* wants to do *x* and that *S* thinks that the pleasure of doing *x* is worth any drawbacks he sees. A case where someone does *x* wanting to do *x*, where he does not do *x* unwillingly or against his will, and where he thinks that doing *x* will have results which are worth any bad consequences he foresees, is precisely the sort of case which one would want to contrast with cases of being forced to do *x*—either by pain, strong emotion, compulsive craving, etc. Doing *x* for pleasure is a paradigm of doing something and *not* being forced to do it. It is in this sense that I call it a “free” action.

Someone who did *x* for pleasure on a particular occasion cannot be said to have been forced to do *x* on that occasion—at least not in any sense of “forced” which could be taken to imply that the agent did *x* against his will, or unwillingly, or involuntarily, or contrary to what he would rather do, or in the absence of any desire to do *x*. In this

respect, an instance of doing *x* for pleasure is an instance of doing *x* freely. A consequence of this is that one cannot consistently deny that there are any free acts in this sense while espousing psychological hedonism or any other doctrine which implies that any acts are done for pleasure. If one embraces such a version of determinism, one must give up pleasure as a concept for explaining action.

The same features which guarantee that an act done for pleasure will be a free act also rule out the possibility of certain (though not all) excuses and justifications should the agent be required to defend his action. Should someone do *x* just for pleasure and in so doing do something wrong or untoward, he cannot honestly defend himself by claiming that he did *x* because it was urgent or imperative that he do *x*. He cannot claim that he did *x* to fulfill an important need or obligation, nor can he claim that he was in any way forced to do *x*. When this is considered together with the point that neglecting certain matters to do things for pleasure is itself a species of wrong action, it is apparent that one who acts for pleasure thereby places himself in a position which might be described as one of relative “moral vulnerability.” This is probably connected with the tendency on the part of certain moralists to regard pleasure as an evil. It also, I strongly suspect, contributes to our intuition that the theory that makes pleasure the end of all action or the sole criterion of what is desirable is indeed “a doctrine worthy only of swine.”

University of Illinois

VIII. FICTIONAL EXISTENCE

CHARLES CRITTENDEN

APPARENTLY we refer to objects which do not exist: we speak about Sherlock Holmes and the Easter Bunny, but never expect to see them in the street or to read about their activities in the newspapers. Some, like Meinong, have believed that although these things do not exist they still have being in some sense. Russell, though once attracted to this view, felt that it led to absurd consequences and so proposed his Theory of Descriptions, the point of which is the contention "that there are no unreal individuals."¹ To support this he offers an analysis of denoting expressions whereby all assertions containing them are analyzed into assertions without referring expressions. Thus appearances are misleading and we do not really refer to fictitious entities after all. Others have followed Russell's lead though rejecting the Theory of Descriptions. Strawson, for example, once held that referring phrases in fiction have a "spurious" use: the author of a work of fiction is not actually referring by using such phrases, because they do not then have their normal, "genuine" function.² Later, however, he revised his opinion and said that denoting expressions in fiction have a "secondary" use: they refer but presumably not in the central, primary way.³ I wish to discuss this issue in terms of what seems to be a crucial distinction. In its light a number of the difficulties disappear.

I

The distinction I want to emphasize is between statements made by an author in *writing* a fictional piece—those which occur in the body of the fiction and which one reads in reading the piece—on the one hand, and statements not in a novel or story but made about fictitious characters in a novel already written, on the other.⁴ It is clear that this is a real distinction. For an assertion of the latter

sort can be true or false in the simplest sense of these words. If one were asked on an examination whether "Hector was a friend of Achilles" was true or false, he would lose credit for marking "true." It is just not true that, in the *Iliad*, Achilles and Hector were friends; if anyone said so he would be uttering a falsehood. On the other hand consider an assertion like the following, taken from the *Iliad* (Book II) and referring to Thersites: "His head ran up to a point, but there was little hair on the top of it." What would establish that this was false? Even if the *Iliad* should later say that Thersites' head was round, and the roundness of his head should play an important part in the narrative, we could not simply count the earlier statement false. One might think that there were two Thersites, or that the head had changed, or something of the sort. The statement is not falsified by a later appearance of its contradictory in the poem. Nor could one verify it by checking it against the facts, for it functions as one of a number of assertions *setting out* the fictional situation. It is made not as reporting a state of affairs existing independently, but rather it is made in the course of *creating* a fictional world. Thus it hardly makes sense to check it against this fictional world, for this would be checking it against itself (at least in part). I shall return to these contentions later.

It is convenient to discuss first assertions about fictitious characters in novels and stories already written. These are true and false in the same way that ordinary empirical assertions are true and false; i.e., to verify them, we must look to the subject matter and note whether the facts are as the statements state. In each case one must first understand the assertion, then have recourse to a separate body of facts, and finally check the assertion against the facts. Of course there is the notable difference that to verify an empirical statement, one

¹ "On Denoting," reprinted in *Logic and Knowledge*, ed. R. C. Marsh (London, 1956), p. 55.

² "On Referring," in *Essays in Conceptual Analysis*, ed. A. Flew (London, 1956), p. 35.

³ *Ibid.*

⁴ This distinction arose out of the oral examination on the dissertation on which this paper is based. I should like to thank my examiners, Professors Max Black, Keith Donellan, and Sydney Shoemaker, for their helpful criticism. Thanks are also due to an anonymous reader for the *American Philosophical Quarterly*, several of whose suggestions I have adopted.

uses his eyes and ears, etc. (or asks someone who has, or reads something written by someone who has, and so on); while to verify a statement about a fictitious being one must read the fictional work in question. To find out whether Hector was the friend of Achilles one must read (parts of) the *Iliad*, or ask someone who has, consult an account, etc. Nonetheless both the empirical assertion and that about fictions are capable of verification; in this way they are logically alike.

A second point of similarity is that both can be about objects. There would be no hesitation in saying that "Achilles killed Hector" was about Achilles, given the appropriate context. (There are problems about "about." But I mean to indicate merely that statements *can* be about fictitious entities; I am not further concerned with the notion.) And it would be admitted too that the person making this assertion had referred to Achilles ("Who is he referring to? Oh, the character in the *Iliad*"). So from an ordinary, unsophisticated point of view, there can be assertions about fictitious characters and references to them. The same conclusions appear if we consider referring as a technical notion, put forward to clarify the role of a certain class of expressions, chiefly nouns, singular pronouns, and definite descriptions. Strawson writes:

If we want to fulfill this purpose [of stating facts about things, persons, and events], we must have some way of forestalling the question, "What (who, which one) are you talking about?" . . . The task of forestalling [this] question is the referring (or identifying) task.⁵

To refer in this sense, then, is to use an expression to identify the object in question for a hearer: to inform him as to which particular thing reference is being made.⁶ A requirement for the successful use of an expression in this way is that the speaker and hearer share knowledge or belief about a certain range of objects; both must have in mind roughly the same set of things for the speaker to be able to identify one of them for his listener. (This is the paradigm situation; it is subject to modification in various ways which I shall not investigate here.) Clearly the use of "Achilles" in "Achilles killed Hector" is to refer in Strawson's sense. For (if the assertion is to be intelligible) both

speaker and hearer must be familiar with the cast of the *Iliad*. On hearing the statement one would know that it was Achilles (not Ulysses or Agamemnon) who did the deed; "Achilles" is used to identify a particular character. This point can be put in another way: if Strawson is right (as I believe he is) about the use which proper names and other expressions have, then these considerations show that names, etc., appearing in assertions about fictitious beings have just the use they have in literal statements. As further evidence parallel logical oddities can be found: "Walter MacFoozleman is tall but there is no one named 'Walter MacFoozleman';" apparently meant seriously, is no more nonsensical than "Flammus was a friend of Achilles, but there is no one named 'Flammus';" spoken in the course of a discussion of the characters in the *Iliad*.

Hence, to sum up the relevant points concerning statements apparently about fictitious characters: these assertions really are about them, both from a nontechnical standpoint, and also from a more sophisticated point of view. Referring phrases (*qua* referring phrases) occurring in them have exactly the same use as in assertions about real objects. Thus, as far as the rules of everyday speech go, Russell is mistaken when he denies that names in assertions about fictitious and mythical beings are actually names.⁷ But the discussion so far has left us with the original problem: if there are fictitious objects, what sorts of thing are they? Evidently (being fictitious and *ipso facto* unreal) they do not exist: do they therefore inhabit the Meinongian realm of Being?

To begin to answer this question, we should turn to the other set of assertions I have mentioned but have so far not discussed, those appearing in the fiction itself. When an author writes a story, he is not reporting any antecedently existing state of affairs, but rather creating a fictitious situation. As Margaret MacDonald has said:

I want to stress this fact that in fiction language is used to create. For it is this which chiefly differentiates it from factual statement. A storyteller performs; he does not—or not primarily—inform or misinform. To tell a story is to originate, not to report.⁸

A fictional work sets out an imaginary world in-

⁵ Strawson, *op. cit.*, p. 40.

⁶ He elaborates this in *Individuals* (London, 1959), pp. 15-17; and in "Critical Notice of *Modes of Referring* by D. S. Shwayder," *Mind*, vol. 71 (1962), pp. 252-254.

⁷ Russell, *op. cit.*, p. 54. Whether he would be disturbed by this consideration is another matter.

⁸ "The Language of Fiction," reprinted in *Philosophy Looks at the Arts*, ed. J. Margolis (New York, 1962), p. 190.

vented (not discovered) by its author. In general this consists of a set of circumstances, characters and objects, on the one hand, and on the other the events in which these things have a part. A sentence in a novel has the function of contributing to the construction of this world, either in introducing the characters and circumstances or in indicating what happens in the story, or perhaps both. These purposes are not of course exclusive.

Of particular importance is the matter of how a character or object comes to be "in the story." There are devices having the function of presenting a character: on reading "Once upon a time there was a bear who . . .," one knows that he is reading a story about a bear, or at least a story in which it is likely that a bear has an important part. But this expression (which after all occurs chiefly in children's literature) is not necessary. Usually an author will put a character into a novel simply by mentioning him: if in the course of reading a novel we should come across the sentence "Suddenly the door burst open and there stood George Wilburfizzle," and this is straightforward narration (not, for example, uttered by one of the persons in the story), we would understand that in the novel there is a character named "George Wilburfizzle." Similarly for other sorts of objects: to place a house, a town, a horse in a story the author simply writes about these things. As a rule there is more to it than this: a fictitious object or person will be characterized in certain ways: George Wilburfizzle will be a wild-eyed man of forty-five, just escaped from an asylum, or perhaps a repentant prodigal son returned from his wanderings. A fictitious being will usually have a *role*; there will be events in which it has a part. But the important point is that something is said to be "in a story" when there are in the narration expressions referring to it and statements about it.⁹ Locutions of these sorts *establish* that the entity is in the tale. If a person were to read the *Odyssey* and deny that there were therein a person named "Odysseus" or a cyclops, we would suspect that he did not understand what is meant by "in the *Odyssey*." To say that " x is in a story," or "there is (exists) a ϕ in a story," then, is to say that there are in the narrative sentences containing " x " and "the ϕ ," "a ϕ ," etc. The author will have used these or similar expressions in the course of telling the story, i.e., in the course of constructing his imaginary world. In so doing he will have created the entities he mentions and

he will have given them their attributes and their history. Fictitious objects are nothing more mysterious than the things written or spoken about when language is used in this way.

It may seem that to use a referring expression in introducing a character into a fictional world is a strange use. For normally, in literal speech, before one can refer to anything he must ascertain that that thing exists. Only then can he refer. Existence precedes reference, so to speak. Yet when one keeps in mind the creative, constructive function of an author's sentences when he is writing fiction, this "introducing" use of denoting expressions is not so odd. That reference in the one case *establishes* existence rather than presupposes it (as in the other) simply brings out the difference between fictional and literal discourse. It may well be that one could not have mastered fictional uses of language unless he had mastered literal discourse, and that in this sense fiction is logically posterior to literal speech. But it does not follow that a reference in fiction is any *less* of a reference; it is not secondary in this way. And of course what is established in fiction is the *fictional* existence of something; no real existence could be so brought about. Whether something "really" exists depends on whether it falls within the same space and time that we occupy, not on language; and questions about real things are settled ultimately by observation, not by noting whether certain expressions occur in certain kinds of literature. These points bring out the difference between fictional and "real" existence and explain why there is nothing contradictory in a statement like "There are things which do not really exist," which might be made about unicorns.

The two sorts of assertion I have emphasized are related as the notes a musician writes down are related to a commentary on his composition. An author constructs an imaginary world by using language. Through the statements he makes while writing or telling a story he creates a group of characters and gives each of them a role. Yet once set out this fictional realm can be described, investigated, discussed, commented on, and the like, just as any subject matter can. Though both the author and his critic (unlike the musician) are making assertions, it is important to see that the statements made by the one have a very different logical character from those made by the other. The notions of truth, reference, and verification do

⁹ Of course there are hard cases; sometimes there are things in stories which are not directly referred to, and there may be other variations. I am setting out the standard situation.

not apply to them in the same way at all; hence the correct explanation of discourse about fictitious beings cannot be as simple as might be thought.

II

In this concluding section I want to show how my view applies to two troublesome examples and to defend it against a widely used argument. The first example is the venerable "The King of France is bald." I hold that what is presupposed depends on the context in which the statement appears. If it is meant literally, spoken, say, by an Englishman in 1700, then the presupposition is that there really exists a King of France. But if it is intended as a remark about a character in Steinbeck's *Pippin IV*, a novel based on the supposition that contemporary France suddenly becomes a monarchy, then the reference presupposes that there is a King of France in the novel. This is so whether the sentence appears in the novel, or whether it does not appear there, but is used in making an assertion about the character in the novel. Actual existence is not presupposed in either instance. However, there is a third case. Suppose that a person utters "The King of France is bald" in a context where there is no King of France at all: two friends are walking along silently, and suddenly one says "The King of France is bald." His companion would naturally think his friend is day-dreaming or giving utterance to some sudden realization; he would try to find some plausible context where there is a King of France. He might even ask the speaker just what King of France he was talking about. If it can be established that the speaker had in mind no context containing a King of France then he would have been simply uttering words and not stating anything true or false. His utterance is meaningful in the sense that the words making it up are meaningful and in that the sentence is grammatically well-formed and could be used in appropriate circumstances for saying something, but as it stands it is not meaningful in the central sense of communicating information. It has not achieved a truth value, so to speak.

Similar but more complicated is the example given by Quine, "The round square cupola on Berkeley College is pink."¹⁰ This statement, if

literally meant, cannot be true or false, for what would be presupposed is that there exists a round square cupola on Berkeley College and no really existing thing can be round and square. Yet this statement can have a truth value when made by someone speaking about a story. Imagine a story which begins: "Berkeley College was built in the dim past, before anyone knew anything about geometry. Its architecture is unique in that it has a round, square cupola, a great attraction for tourists today. . . ." Now the criterion for saying whether something is in a story is what the story says: imaginary writing is limited not by facts about the real world or even, with some qualifications, by logic but only by the author's imagination.¹¹ If a piece of logical science fiction mentions something contradictory and gives it a role, then this object is in the story. So a statement about such a thing is true if in the story the object has the characteristic asserted of it, while if the object does not have the characteristic the statement is false. Taken in other contexts, Quine's sentence is like "The King of France is bald": if it appears in a story about an imaginary round square then it has no truth value but presupposes a round square in the story. Finally, if there is no imaginary round square where it occurs the same considerations apply to it as to "The King of France is bald" in similar circumstances.

Next I shall discuss an objection which has often been raised against the sort of view I propose. The distinction drawn above between entities which are real and existing, and those which occur only in fiction, is not unlike the distinction between *existence* and *being* introduced by Meinong and used, and later abandoned, by Russell. Russell once wrote:

Numbers, the Homeric gods, relations chimeras, and four-dimensional spaces all have being, for if they were not entities of a kind, we could make no propositions about them. Thus being is a general attribute of everything, and to mention anything is to show that it is. Existence, on the contrary, is the prerogative of some only amongst beings.¹²

There is a question as to how these remarks ought to be interpreted. They could be taken as expressing the view I have suggested. But this was probably not their intended meaning, and the usual

¹⁰ "On What There Is," in *From a Logical Point of View* (Cambridge, 1953), p. 5.

¹¹ Compare this with what Malcolm says about dreams: "That something is implausible or impossible does not go to show that I did not dream it. In a dream I can do the impossible in every sense of the word. I can climb Everest without oxygen and I can square the circle." *Dreaming* (London, 1962), p. 57.

¹² *The Principles of Mathematics*, 2d ed. (New York, 1938), p. 449.

interpretation is reflected in Quine's contention that the distinction between existence and being (or "subsistence," as Quine's straw man has it) leads to an "overpopulated universe" which "offends the sense of us who have a taste for desert landscapes." Furthermore, the realm of subsistence "is a breeding ground for disorderly elements."¹³ How are we to decide how many entities having subsistence but not existence there are at any given place? How can we tell when two subsistent objects are the same or are different? In short, Quine is arguing, there are no criteria for deciding when there is such an entity or for distinguishing one from another. These contentions may have force against the doctrines actually held by Meinong and by Russell, but they have none against my own. In the first place, admitting that fictional and other imaginary objects are actual entities is not objectionable—in fact, if I am right, we could not understand novels or plays, or indeed accounts of dreams, mythology, and the like without conceiving imaginary beings as genuine objects. Nor is there any reason for not continuing these practices, as long as the differences between imaginary beings and real, existing entities is kept in mind. Fictional characters are not shadowy creatures half-existing in an ontological limbo; they are the familiar things we read about and tell stories about but never actually meet on the face of the earth. Likewise Quine's question about criteria for what imaginary beings there are and for distinguishing one from another is answered by pointing out just the criteria we make use of in understanding fiction. To tell whether an object appears in a novel we must notice whether it is mentioned by the author in his narrative, as I have argued. And we can find out when it is the same thing as something else mentioned by noticing whether it is *spoken of* as the same. No one who comes across the sentence "Mr. Jones looked intently at his cat" in a story is going to confuse the fictitious Mr. Jones

with his fictitious cat. Of course there could be a story where Mr. Jones did turn into a cat: a reader would be informed of this by the occurrence of locations such as "woke up one morning to find he had become a cat." The criteria for similarity concerning imaginary objects are found in the account, I contend. But the point is that there *are* criteria; in fact, the practice of writing fictional works depends on them. This would not be an intelligible use of language unless it included some way of telling when two things are the same or are alike. Quine's objection overlooks criteria actually in use.

Now I want to pass to a different matter. In the latter part of his essay Quine speaks of the problem of "accepting an ontology." "Our acceptance of an ontology is," he says, "similar in principle to our acceptance of a scientific theory, say a system of physics: we adopt, at least insofar as we are reasonable, the simplest conceptual scheme into which the disordered fragments of raw experience can be fitted and arranged."¹⁴ And presumably a conceptual scheme is simpler if it has no commitment to non-existent things. (Perhaps Quine has this sort of conceptual system in mind when he speaks of his "taste for desert landscapes.") I do not wish to raise the issue of the value of this sort of system-construction; I wish to point out only that it is a very different enterprise from that of setting out the rules in force for actual linguistic practice and indicating their ontological commitments. Quine evidently supposes that the "simplest" conceptual scheme would not allow references to non-existent entities—but even if this is granted, it has no bearing on the matter of whether we do in fact refer to them in our non-theoretical moments.

My contention throughout has been that we do make such references and that this has no unfortunate logical consequences, for the apparent difficulties disappear when we look closely at actual usage.

Florida State University

¹³ Quine, *op. cit.*, p. 4.

¹⁴ *Ibid.*, p. 16.